ANOMALY DETECTION IN REAL-WORLD TEMPORAL NETWORKS

Pablo Moriano Salazar

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the School of Informatics, Computing, and Engineering,
Indiana University

May 2019

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

L. Jean Camp, Ph.D., Chair

Yong-Yeol Ahn, Ph.D.

Filippo Radicchi, Ph.D.

Raquel Hill, Ph.D.

April 9, 2019

Esta disertación esta dedicada a mis padres, Adiela y Jorge, y a mi esposa Claudia.

Su amor ha sido y será lo mas lindo de mi vida.

Acknowledgments

Foremost I have to thank GOD. He has been my best friend during this journey. He has giving me life, health, wisdom, and courage to reach this point. He has always be with me although I do not deserve it. Thanks to his infinite mercy and grace, I was fortunate to be able to complete my doctoral studies in the United States, just as I dreamed one day. I hope that through the gift that I received, I will be able to use and serve one another as a good steward of God's varied grace.

I do not even know how to thank my lovely wife, Claudia Castro. This dissertation would not have been possible without her unending support. She left all what she loved in Colombia to come to the United States with me, so I could live my dream. She has been always there for me and have loved me through my absences, frustration, and grumpiness. Without her, I would not have succeeded. I hope we will remember all these years forever, together. I love you so much ♡.

I am infinitely grateful to my parents, Adiela and Jorge. They taught me the more important lessons in life. Those that you never are going to learn even having access to the best worldwide education. They have been the best example of discipline, honesty, humility, and mercy. Thanks to their prayers, I was able to finish this dissertation despite the multiple ups and downs during the process. Thank you mum and dad for letting me be the person I want to be. They never questioned my decision to move abroad to pursue my doctoral studies and have been my support for everything.

I am also thankful to a number of people who directly or indirectly have helped or supported me during my doctoral studies. My committee members deserve individual acknowledgements.

I owe the greatest gratitude to my advisor, Jean Camp. Jean introduced to me the fascinating world of security research. She has been an amazing source of support and guidance. I appreciate all the time she spent discussing about research, editing my papers, and showing me how to treat people with respect and delicacy. Jean also guided me to find a community of researchers to collaborate outside Indiana University. It was an honor for me to work with her.

I would like to thank Yong-Yeol Ahn for teaching me how to do rigorously scientific research. YY is an authentic super critical scientist who can able to connect points from multiple disciplines

to provide insights from different angles. I have not seen in many such extraordinary ability. I really enjoyed the process of working with him in the event detection project that emerged from a class assignment.

I would like to thank Filippo Radicchi for agreeing being my first advisor during my Ph.D. journey. Filippo was always open to hear about my progress, problems, and effectively provide me support at different stages. Although the project that we worked together at the beginning did not pan out, I learned from him about how to code in Python and write more efficient algorithms. I really appreciate the time he spent with me on the whiteboard doing mathematical modeling of networks and about probability concepts. I think that he is one of the best doing that.

I am also very grateful to Raquel Hill. Raquel helped me to disentangle the BGP anomaly detection project. She carefully listened to me about my progress and provided me key insights to arrive at the final results. I will remember the exciting conversations we had in her office about routing and in general about how the Internet works. She was also an outstanding lecturer in the Security for Networked Systems class. That class definitely was one of the best classes I took during my Ph.D.

I was lucky to work with the Advanced Security Research Group at Cisco during my summer internships. It was a great source of inspiration to work on different projects and people in an industry environment. In particular, I would like to thank Steven Rich for being an excellent mentor and collaborator. Steve introduced me to the insider threat problem and helped me to find all the required resources to get that project done. I appreciate that Steve was always open to integrate scientific research knowledge in applied security problems. Within Cisco, I also would like to thank Jim Warren, Yousef Iskander, Jared Pendleton, and Robert Broberg, for assuring that my internship experiences were fruitful.

Prior my doctoral studies, I had a great mentor during my masters. I would like to thank Jorge Finke for that. Jorge had a tremendous positive impact on my career by teaching me how to reason about problems and introducing me to the field of network science. He always encouraged me to pursue my doctoral studies in the United Sates, and I am very grateful for have taken that advice.

Despite professional guidance, more importantly, Jorge offered me his friendship and helped to navigate through two very difficult academic situations I had to face during my time at Indiana University.

I would like also to thank Katy and John Sparks for being our American parents during our stay in Bloomington. I appreciate their sincere friendship and for welcome Claudia and I across different situations that made our life in Indiana more joyful. Thank you Katy and John for emotionally supporting us and for the amazing board game sessions during Thanksgiving days. I will remember them forever ☺.

I appreciate the friendly environment in the HATS research group. Over the years, we had several moments of joy and support for the exchange of ideas. I enjoyed our weekly meetings and board game sessions with (not in any particular order) Jacob Abbot, Behnood Momenzadeh, Jayati Dev, Joshua Streiff, as well as all other members of the group. Similarly, from the CNetS group, I would like to thank Jaehyuk Park and Alexander Barron for the insightful conversations we had during these years and for being good neighbors. I would also like to thank Tim Kelley for the collegial act of sharing the necessary templates for this dissertation.

I want also thank Beverly Diekhoff for providing amazing administrative help to navigate through this process. My special thanks to Bruce Shei for providing me support on setting up computer systems to run computational experiments and Kenneth Bikoff who patiently edited several of my papers and this dissertation. Finally, I thank David Nemer for helping me to find a place to finish the writing of this dissertation in the libraries of the University of Kentucky in Lexington.

* This dissertation was supported in part by NSF CISE #1565375, Cisco Research #591000, and Google Privacy & Security Focused Research. I thank the Colombian Administrative Department of Science, Technology, and Innovation (COLCIENCIAS) through the "Programa de formación doctoral en el exterior" (call 617 of 2013) and COLFUTURO through the "Programa Crédito-Beca 2013" for their support. I would also like to thank the School of Informatics, Computing, and Engineering at Indiana University and supporters of student travel funding that I received during the past six years.

Pablo Moriano Salazar
ANOMALY DETECTION IN REAL-WORLD TEMPORAL NETWORKS

Detection of anomalous events relies on the collection, filtration, and analysis of diverse types of temporal data. Interactions derived from such data can be modeled as networks to provide a better understanding of the structure and dynamics of the underlying systems. This dissertation examines the temporal evolution of Internet-scale phenomena to provide a more nuanced characterization of normal functionality and anomalous behavior—usually undesired and often malicious.
Characterizing regular behavior is often a prerequisite for identifying these anomalies. However, the volume and patterns of interactions during a system's evolution under particular circumstances may be highly variant. In security, I create hypotheses about the nature of attacks as a core component of detection. In the insider threat, I hypothesized that malicious insiders would require access to identifiably more diverse repositories than the non-malicious. In routing, my first hypothesis was that the role of nation-state actors could be identified using traditional macroeconomic analyses. My second hypothesis, that control plane hijacking could be identified by leveraging k-shell decomposition of Autonomous System level graphs, could not be validated. In contrast, the analysis of inter-arrival times of route announcements provided clear identification and early warning of large-scale incidents.
This dissertation contributes to a more comprehensive understanding of security threats using data and network science methods. Specifically, I use (i) Graph mining to show that surprising patterns about community structure and k-shell decomposition of graphs can be leveraged to detect classes of anomalies. Leveraging (ii) Graph robustness, I show how community detection-based methods are less biased against the density of edges in the system, providing a robust approach to detect anomalous behavior. Finally, I illustrate the potential of (iii) Graph anomaly detection for identifying anomalies in different real-world scenarios, including (a) email interactions, (b) social media, (c) code repositories, and (d) Internet control-plane updates.

L. Jean Camp, Ph.D., Chair

Yong-Yeol Ahn, Ph.D.

Filippo Radicchi, Ph.D.

Raquel Hill, Ph.D.

Contents

List of Figures xiii

List of Tables xviii

Chapter 1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Thesis Statement . . . 4
- 1.3 Research Questions and Overview . . . 4
  - 1.3.1 Part I: Anomaly Detection in the Society . . . 5
    - 1.3.1.1 Chapter 3: Community-Based Event Detection in Temporal Networks . . . 6
  - 1.3.2 Part II: Anomaly Detection in an Organization . . . 7
  - 1.3.3 Chapter 4: Protecting Organizational Assets Through Graph Mining . . . 7
  - 1.3.4 Part III: Anomaly Detection on the Internet . . . 8
  - 1.3.5 Chapter 5: Macroeconomic Analysis of Routing Anomalies . . . 10
  - 1.3.6 Chapter 6: Characterization of Internet Routing Anomalies Through Graph Mining . . . 11
  - 1.3.7 Chapter 7: Bursty Announcements for Early Detection of BGP Routing Anomalies . . . 13
- 1.4 Research Impact . . . 14
- 1.5 Outline . . . 14

Chapter 2 Related Work 16
- 2.1 Security as Anomaly Detection . . . 16
- 2.2 Graph-Based Anomaly Detection . . . 17
  - 2.2.1 Graph-Based Event Detection . . . 18
  - 2.2.2 Community Structure . . . 22
- 2.3 BGP . . . 24
  - 2.3.1 IP Prefixes . . . 24
  - 2.3.2 AS Numbers . . . 25
  - 2.3.3 Prefix Hijacking . . . 25
  - 2.3.4 Detection of BGP Anomalies . . . 25
  - 2.3.5 Mitigation of BGP Anomalies . . . 27
  - 2.3.6 Macroeconomics of Security . . . 29
  - 2.3.7 Political Operations . . . 30
- 2.4 Insider Threats as Anomalies . . . 32

2.4.1 Characterization of Insider Threats . . . 32
2.4.2 Insider Threat Detection Using Graph-Based Methods . . . 33

Chapter 3 Community-Based Event Detection in Temporal Networks 35
3.1 Introduction . . . 35
3.2 Problem . . . 35
3.3 Methods . . . 37
3.3.1 Data . . . 37
3.3.1.1 Enron Email Communication Network . . . 37
3.3.1.2 Twitter Interaction Networks During the Boston Marathon Bombing . . . 38
3.3.2 Network Representation . . . 38
3.3.2.1 Detection Problem . . . 38
3.3.3 Algorithm Evaluation . . . 39
3.3.4 The Proposed Detection Algorithm . . . 40
3.3.5 Performance Benchmark . . . 42
3.4 Results . . . 43
3.4.1 Enron . . . 43
3.4.2 Boston Marathon . . . 45
3.4.2.1 Mention Network . . . 45
3.4.2.2 Retweet Network . . . 48
3.5 Conclusion . . . 50

Chapter 4 Insider Threat Modeling 52
4.1 Introduction . . . 52
4.2 Problem . . . 52
4.3 Methods . . . 55
4.3.1 Temporal Abstraction . . . 55
4.3.2 Bipartite Graph Abstraction . . . 56
4.3.3 One-Mode Projection Abstraction . . . 56
4.3.4 Detection Problem . . . 57
4.3.5 Algorithm Performance Abstraction . . . 59
4.3.6 Algorithm Performance Measure . . . 60
4.3.7 Proposed Algorithm . . . 61
4.3.8 Dataset . . . 64
4.4 Results . . . 66
4.4.1 Bipartite Graph Properties Series . . . 67
4.4.2 One-Mode Projection Graph Properties series . . . 68
4.4.3 Algorithm Evaluation . . . 70
4.4.4 Algorithm Performance . . . 72
4.5 Conclusion . . . 75

Chapter 5 Macroeconomics of Routing Anomalies 80
5.1 Introduction . . . 80
5.2 Problem . . . 81

5.3 Methods . . . 82
5.3.1 Data Sources . . . 83
5.3.1.1 Independent Variables . . . 83
5.3.1.2 Dependent Variable . . . 85
5.3.1.3 Data Preprocessing . . . 87
5.3.2 Statistical Model . . . 88
5.4 Results . . . 90
5.4.1 Anomaly Time Series . . . 90
5.4.2 Anomaly Distribution . . . 91
5.4.3 Regression Analysis . . . 93
5.4.4 Cluster Analysis . . . 97
5.5 Conclusion . . . 99

Chapter 6 Characterizing Routing Anomalies Through Graph Mining 105
6.1 Introduction . . . 105
6.2 Problem . . . 105
6.3 Methods . . . 107
6.3.1 Data Sources . . . 107
6.3.1.1 Routing Anomalous Events . . . 108
6.3.1.2 BGP Data . . . 109
6.3.1.3 AS-Level Graph Representation . . . 110
6.3.1.4 Topological Properties . . . 112
6.4 Results . . . 116
6.4.1 Global Structure Measures . . . 117
6.4.2 Path Length . . . 120
6.4.3 Community Structure . . . 123
6.5 Conclusion . . . 126

Chapter 7 Bursty Announcements for Early Detection of BGP Routing Anomalies 130
7.1 Introduction . . . 130
7.2 Problem . . . 130
7.3 Methods . . . 132
7.3.1 Data Sources . . . 133
7.3.1.1 BGP Data . . . 133
7.3.1.2 Routing Anomalous Events . . . 133
7.3.2 Burstiness of Announcements . . . 135
7.3.3 Detection Method . . . 135
7.4 Results . . . 136
7.4.1 Feeder Contribution Analysis . . . 138
7.4.2 Collectors' Disruption Perception . . . 140
7.4.3 Inter-Arrival Time Analysis . . . 142
7.4.4 Anomaly Detection . . . 148
7.5 Conclusion . . . 151

Chapter 8 Conclusions 155

8.1 Summary of Contributions . . . 155
8.1.1 Chapter 3: Community-Based Event Detection in Temporal Networks . . 155
8.1.2 Chapter 4: Insider Threat Modeling . . . 156
8.1.3 Chapter 5: Macroeconomics of Routing Anomalies . . . 157
8.1.4 Chapter 6: Characterizing Routing Anomalies Through Graph Mining . . 158
8.1.5 Chapter 7: Bursty Announcements for Early Detection of BGP Routing Anomalies . . . 158
8.2 Future Work . . . 159
8.2.1 Limits of Community-Based Event Detection . . . 159
8.2.2 Understanding Software Quality in Developer-Component Temporal Graphs 160
8.2.3 Relationship Between On-Line and Off-Line Cross Country Conflicts . . . 160
8.3 Final Remarks . . . 161

Bibliography **163**

Appendix A Additional Topological Properties BGP Graph Mining **190**
A.1 Global Structure . . . 190
A.1.1 An Indonesian ISP Hijacking the World . . . 190
A.1.2 Global Collateral Damage of Telecom Malaysia Leak . . . 191
A.1.3 Large Scale BGP Hijack in India . . . 192
A.2 Community Structure . . . 193
A.2.1 An Indonesian ISP Hijacking the World . . . 193
A.2.2 Global Collateral Damage of Telecom Malaysia Leak . . . 197
A.2.3 Large Scale BGP Hijack in India . . . 197

Appendix B Remaining Collectors BGP Burstiness Analysis **200**
B.1 Collectors' Disruption Perception . . . 200
B.2 Inter-Arrival Time Analysis . . . 201
B.3 Anomaly Detection . . . 201

Curriculum Vitae

List of Figures

1.1 Schematic representation of the proposed event detection method. Both networks have the same communities but different patterns of communication within and across them. (A) When there is no global event, most communication occurs within communities. (B) When a global event occurs, more communication may happen across communities because of the global relevance of the event and because of its virality. 6

1.2 Malicious activity in the bipartite graph. The top panel represents an unusual interaction between Bob and Eve with component "E." The bottom panel represents the corresponding one mode projection graph with the anomalous edge crossing communities. Please refer to Figure 1.1 for a more abstract schematic illustrating the method. 9

1.3 Geographic location of the countries with considerable high number of anomaly/ASes and poor governmental practices during the observation period. 11

1.4 A schematic representation of the graph under k-shell decomposition for different values of shell k_s. The AS-level graph under normal conditions is on the left and the same topology when a BGP hijack is committed is on the right. The hijacker is highlighted with the black hat. 12

1.5 Difference between events produced a non-bursty sequence (above) and a bursty one (below). 14

2.1 A graph with three communities enclosed by the dashed circles. 22

3.1 Time series of Enron events. (a) Time series of the number of emails. (b) Time series of the difference between the inter- and intra-community link ratios. (c) Time series of the number of emails classified by topics. (d) Time series of the difference between the inter- and intra-community link ratio classified by topics. 45

3.2 Performance comparison for the Enron case when $m = 2$ weeks. (a) ROC. (b) PRC. . 46

3.3 Performance comparison for the Enron case when $m = 7$ weeks. (a) ROC. (b) PRC. . 46

3.4 Time series analysis of the mention network. (a) Time series of the number of mentions. (b) Time series of the difference between the inter- and intra-community link ratios. (c) Distribution of the number of hashtags based on the total number of links (horizontal axis) and the difference of inter- and intra-community links (vertical axis) during the interval 14:00-16:00 EST on 2013-04-15. (d) Same as (c) during the interval 16:00-18:00 EST. (e) Same as (c) during the interval 18:00-20:00 EST. Hashtags related to the Boston Marathon bombing are highlighted. 48

3.5 Time series analysis of the retweet network. We perform the same analysis as for the mention network and report the number of retweets in (a), the difference between inter- and intra-community link ratios in (b), the distribution of the number of hashtags based on the total number of links (horizontal axis); and the difference of inter- and intra-community links (vertical axis) during the interval 14:00-16:00 EST on 2013-04-15 in (c). (d) Same as (c) during the interval 16:00-18:00 EST. (e) Same as (c) during the interval 18:00-20:00 EST. Hashtags related to the Boston Marathon bombing are highlighted. 49

4.1 Bipartite graph abstraction. The top panel represents the engineer projection. The middle panel represents the original bipartite graph. The bottom panel represents the component projection. 57
4.2 Abstraction of the detection problem. The top panel refers to the sequence of intervals that are used to build the graphs (here the graph formation interval $\mathfrak{m} = 1$). The bottom panel illustrates the aggregation of intervals to evaluate the performance of the detection algorithm (here detection resolution $l\mathfrak{m} = 2$). The vertical arrows represent the location of an anomalous event in both temporal representations. The horizontal arrows illustrate the sets E and $\hat{\mathsf{E}}$. 60
4.3 Correspondence between the graph formation intervals and the intervals at which the algorithm is evaluated. Here $\mathfrak{m}$ is the number of intervals at which the algorithm is evaluated. 61
4.4 Malicious activity as identified in the bipartite graph. The top panel represents an unusual interaction between Bob and Eve with component "E." The bottom panel represents the corresponding one mode projection graph with the anomalous edge crossing communities. 63
4.5 Time series of the number of nodes (top panel), edges (middle panel), and connected components (bottom panel) for the bipartite graphs. 68
4.6 Time series of the avg. degree (top panel), max. degree (middle panel), and max. weight (bottom panel) for the bipartite graphs. 69
4.7 Time series of the number of nodes (top panel), edges (middle panel), and connected components (bottom panel) for the one-mode projection graphs. 70
4.8 Time series of the avg. degree (top panel), max. degree (middle panel), and max. weight (bottom panel) for the one-mode projection graphs. 71
4.9 Time series of the number of nodes (top panel), edges (middle panel), and components for the one-mode projection graphs (bottom panel). 72
4.10 Time series of the avg. degree (top panel), max. degree (middle panel), and max. weight (bottom panel) for the one-mode projection graphs. 73
4.11 Time series of the intra- minus inter-community edge ratio for the one-mode projection graphs. 73
4.12 Algorithm performance for detection resolution $4\mathfrak{m}$. Results of the proposed method are comparable with the ones of the random approach. 75
4.13 Algorithm performance for detection resolution $8\mathfrak{m}$. Results of the proposed method begin being better than the ones of the random approach. 75
4.14 Algorithm performance for detection resolution $16\mathfrak{m}$. Results of the proposed method begin being much better than the ones of the random approach. 76

4.15 Algorithm performance for detection resolution 26m. We obtain about 92% of F1 score. 76

5.1 Daily number of anomalies. The dashed line represents the unweighted LOESS fit. The blue band corresponds to the 95% confidence interval of the unweighted LOESS fit. 91
5.2 CCDF of anomalies originated per country. 93
5.3 Anomalies per year for countries that are in the upper 2.5% tail of the anomaly distribution. 94
5.4 Anomalies vs. SIS per year per country. 95
5.5 Distribution of the number of SIS per year for countries that are in the upper 2.5% tail of the SIS distribution. 96
5.6 SIS vs. WGI for all countries over all years. 97
5.7 Anomaly/ASes ratio in the SIS versus WGI plane for 2011. 99

6.1 Graphical representation of the AS-level graph. 115
6.2 A schematic representation of the graph under k-shell decomposition for different values of shell k_s. The AS-level graph under normal conditions is on the left and the same topology when a BGP hijack is committed is on the right. The hijacker is highlighted with the black hat. 116
6.3 Maximum degree Indonesia event. 118
6.4 Nodes per crust Indonesia event. 118
6.5 Maximum degree Malaysia event. 119
6.6 Nodes per crust Malaysia event. 119
6.7 Maximum degree India event. 120
6.8 Nodes per crust India event. 121
6.9 Average path length in the crust Indonesia event. 122
6.10 Average path length in the crust Malaysia event. 122
6.11 Average path length in the crust India event. 123
6.12 Clustering per crust Indonesia event. 124
6.13 Average size components crust Indonesia event. 125
6.14 Clustering per crust Malaysia event. 125
6.15 Average size components crust Malaysia event. 126
6.16 Clustering per crust India event. 127
6.17 Average size components crust India event. 127

7.1 Time series of the number of routers peering with collectors. Collectors are ordered in alphabetical order. Major ticks correspond to nine-month intervals while minor ticks correspond to one-month intervals. 139
7.2 Time series of the number of announcements from AS 4761 that collectors received before, during, and after the Indosat incident in 2014 for the top four collectors. Major ticks correspond to six-hour intervals while minor ticks correspond to two-hour intervals. 141
7.3 Time series of the number of announcements from AS 4788 that collectors received before, during, and after the Telecom Malaysia incident in 2015. 142
7.4 Time series of the number of announcements from AS 9498 that collectors received before, during, and after the Bharti Airtel Ltd. incident in 2015. 143

7.5 Joint distribution based on the the burstiness (horizontal axis) and number of announcements (vertical axis) during one day interval around the Indosat incident. . . . 145
7.6 Monte Carlo test for burstiness. Last column corresponds to the observations of the AS responsible for the incident, AS 4761. The test statistic, the burstiness observed during the interval of the attack, is marked with a cross. 146
7.7 Joint distribution based on the total number of announcements and their burstiness during one day interval around the Telecom Malaysia incident. 147
7.8 Monte Carlo test for burstiness. The last column corresponds to the observations of the AS responsible for the incident, AS 4788. 148
7.9 Joint distribution based on the total number of announcements and their burstiness during the one day interval around the Bharti Airtel Ltd. incident. 149
7.10 Monte Carlo test for burstiness. The last column corresponds to the observations of the AS responsible for the incident, i.e., AS 9498. 150
7.11 $Q_{4761\to B}$ time series for the Indosat incident. 151
7.12 $Q_{4788\to B}$ time series for the Telecom Malaysia incident. 152
7.13 $Q_{9498\to B}$ time series for the Bharti Airtel Ltd. incident. 153

A.1 Number of nodes Indonesia event. 191
A.2 Number of edges Indonesia event. 191
A.3 Nodes per core Indonesia event. 192
A.4 Number of nodes Malaysia event. 193
A.5 Number of edges Malaysia event. 193
A.6 Nodes per core Malaysia event. 194
A.7 Number of nodes India event. 194
A.8 Number of edges India event. 195
A.9 Nodes per core India event. 195
A.10 Clustering coefficient Indonesia event. 196
A.11 Clustering per core Indonesia event. 196
A.12 Clustering coefficient Malaysia event. 197
A.13 Clustering per core Malaysia event. 198
A.14 Clustering coefficient India event. 198
A.15 Clustering per core India event. 199

B.1 Time series of the number of announcements from AS 4761 that collectors received before, during, and after the Indosat incident in 2014 for the top four collectors. Major ticks correspond to six-hour intervals while minor ticks correspond to two-hour intervals. 200
B.2 Time series of the number of announcements from AS 4788 that collectors received before, during, and after the Telecom Malaysia incident in 2015. 201
B.3 Time series of the number of announcements from AS 9498 that collectors received before, during, and after the Bharti Airtel Ltd. incident in 2015. 202
B.4 Joint distribution based on the the burstiness (horizontal axis) and number of announcements (vertical axis) during one day interval around the Indosat incident. . . . 203

B.5 Monte Carlo test for burstiness. Last column corresponds to the observations of the AS responsible for the incident, AS 4761. The test statistic, the burstiness observed during the interval of the attack, is marked with a cross. 204
B.6 Joint distribution based on the total number of announcements and their burstiness during one day interval around the Telecom Malaysia incident. 205
B.7 Monte Carlo test for burstiness. The last column corresponds to the observations of the AS responsible for the incident, i.e., AS 4788. 206
B.8 Joint distribution based on the total number of announcements and their burstiness during the one day interval around the Bharti Airtel Ltd. incident. 207
B.9 Monte Carlo test for burstiness. The last column corresponds to the observations of the AS responsible for the incident, i.e., AS 9498. 208
B.10 $Q_{4761\to B}$ time series for the Indosat incident. 209
B.11 $Q_{4788\to B}$ time series for the Telecom Malaysia incident. 210
B.12 $Q_{9498\to B}$ time series for Bharti Airtel Ltd. 211

List of Tables

3.1 Enron's event description. 43
3.2 Boston Marathon bombing event description. 46

4.1 Summary of precipitating events during the observation period. 65

5.1 Five-dimensional regression model variables. 85
5.2 OLS regression estimates. 95

6.1 Summary of large-scale routing incidents. 109
6.2 Summary of graph topological properties and its relationship with Internet's performance. 112

7.1 Geographical location and date of first dump of collectors. Collectors are ordered in alphabetical order. 137

1 Introduction

"The beginning of knowledge is the discovery of something we do not understand."

— Frank Herbert

1.1 Motivation

Real-world relational data is omnipresent in today's world. Interactions derived from such data can be modeled as networks. (Throughout this dissertation I will use the terms network and graph interchangeably.) The resulting network view can provide a better understanding of the structure and dynamics of the underlying systems. Study of the temporal evolution of these networks is crucial to provide a more detailed characterization of the system's function. Irregularities, i.e., anomalies, in the general evolution of these networks are usually associated with undesired and often critical behavior, in particular, in computer security [Chandola et al., 2009]. Thus, there is a need to test the ability of temporal graph-based methods to contribute to the understanding and mitigation of anomalies in these time-varying scenarios [Akoglu et al., 2015].

Collecting and analyzing large amounts of data has been of great utility in different fields ranging from physics [Vespignani, 2009] to sociology [Lazer et al., 2009], including computer security [Sommer and Paxson, 2010]. Structured analysis using methods from other fields—from economics to complex systems—is only now being widely leveraged in computer security. This dissertation analyzes large-scale relational data in computer security for anomaly detection.

In particular, I develop data-mining and temporal graph-based methods to detect anomalies in time-varying, real-world computer security systems—a field in which anomalies are constantly evolving as intelligent attackers respond to the models of defenders. Temporal graph-based analysis combines the strengths of time series and the power of social network analysis, and has the

potential for automating the detection of previously hidden attacks [Ide and Kashima, 2004, Neil et al., 2013].

The use of temporal graph-based methods as opposed to traditional time series of data points for anomaly detection has numerous advantages. First, graphs allow natural labeling of their elements, e.g., domain names, IP addresses, or port numbers for the nodes; and latency, throughput, or bandwidth for the edges; which is useful to understand long-term correlations. Second, graphs capture the interdependencies between elements, which provides a better understanding of the structure and dynamics of the underlying systems. This is useful to track the interaction between groups of elements. Finally, graphs generate new features that might not be robust in response to normal changes in the system. Graph-based analysis allows us to identify stationary trends in the design of reliable anomaly detection algorithms. In the field of security analytics, temporal graph-based anomaly detection methods can be directly applied to problems of intrusion detection, classification of network traffic, and understanding of massive network outages.

Characterizing regular behavior is often a prerequisite for the identification of anomalies. This is made more difficult because variations in the volume of interactions during a system's evolution may be the norm, or may be normal only under particular circumstances. Identifying stationary trends that enable us to design reliable detection algorithms remains an open challenge. Extensive yet careful data compilation, rigorous analysis, and iterative modeling are all necessary to secure our interconnected society.

For example, consider the problem of detecting large-scale events derived from email communications inside a particular organization or from society using Twitter data. One might expect that during a certain event, the volume of communications tends to increase, e.g., when a specific deadline approaches or when there is an incident of global interest. Based on the nature of these interactions, the events should not be labeled as anomalous, but how can we derive clearer detection signatures for event detection in these types of systems? The analysis of the community interaction between members of communication graphs offers a complementary explanation to detect particular events that is not biased toward the volume of interactions.

In the context of software development, how can we ensure the protection of the integrity of organizational assets? In this example, the complex set of interactions between developers and code repositories again can be captured by time-varying graphs. A particularly important observation is that there are observable external events (precipitating events) that might influence potential risk behavior of employees toward organization assets. Such precipitating events are well documented in organizational literature. As in the previous case, the development of a method that leverages the notion of the community structure in the interactions between developers and code repositories produces a new signature for detection that allows quantitatively verification.

In the Internet control-plane, are there any macroeconomics factors that propitiate the generation of a greater number of routing anomalies[1] from some countries as opposed to others? The use of macroeconomic analysis provides a theoretical grounding to identify those differences. In particular and, in conjunction with multiple linear regression and cluster analysis, it is possible to illustrate that there are key macroeconomic variables that are consistently identifiable as different between countries that produce a disproportionate number of such anomalies. This first analysis provided a foundation for understanding the dynamics and patterns of hijacks, which enabled a hypothesis-driven evaluation of control-plane attacks from a graph-theoretical perspectives.

Internet control-plane large-scale outages due to routing anomalies have been increasing over the last several years [Vervier et al., 2015]. From the technical computer security point of view, there are many proposed cryptographic solutions to mitigate them. Ultimately, they would imply dramatic changes to the Internet infrastructure. Although technically correct, their applicability seems to be infeasible. Spatio temporal analysis of routing anomalies, on the other hand, provides key insights that elucidate the motivation behind the potentially benign increase in these types of incidents. In particular, the use of graph mining to study the structure of Autonomous System (AS)-level graphs reveals the challenges of studying whether anomalies impact the structure of those graphs and its differentiation with regular changes. However, by studying the dynamics

[1]Routing anomalies are those incidents that happen on the control-plane at the Border Gateway Protocol (BGP) level.

of the system instead of its topology, we ask a different question: Are there any changes in the dynamics of route updates when a large-scale BGP incident is incipient? In this case, a quantitative characterization of the system update dynamics based on burstiness allows us to extract a clear distinction between incident and non-incident scenarios.

This type of rigorous analysis provides quantitative validation of apparent common sense explanations. Now, I provide the thesis statement and the problems I address as a part of the contribution of this dissertation.

1.2 Thesis Statement

This dissertation aims to develop data-driven and network science methods to understand the emergence of anomalies in longitudinal security data. In particular, I demonstrate:

1. the use of graph mining methods to better characterize and identify anomalous events in email, social media, and user-system interactions;

2. the potential of spatio temporal analysis for the understanding of hidden factors behind Internet control-plane events; and

3. that such data-intensive hypothesis-driven analyses offer new insights to solve real-world computer security challenges.

I discuss below my main contributions to the anomaly detection problem in the security domain using using data and network-science methods.

1.3 Research Questions and Overview

In Part I, we[2] propose a method for detection of large events in temporal communication networks. In doing so, we analyzed data from email communications in Enron during its collapse and tweets

[2]In this dissertation, I use the pronoun "we" as none of this work could have been successfully completed without collaboration with others.

during the Boston Marathon bombing event. We characterized the interaction between users in both systems using temporal graphs. We show that by leveraging changes in the proportion of inter- and intra-links in the structure of these graphs, it is possible to identify a detection signature around the events of interest.

In Part II, we turn our attention to the complex interactions happening in the process of software development in an organization. In particular, we extrapolate the previous method to provide a signature for anomaly detection in the case of the committing behavior of developers in proprietary code. The main idea relies on capturing committing interactions between developers and code repositories using a bipartite graph. We then use this bipartite graph as a substrate to extract the developer projection and analyze deviation patterns of commits to identify deviations with respect to the commonly contributed repositories. We show that known precipitating events within the organization are correlated with these anomalous distributions of commits.

In Part III, we focus our attention on the Internet—one of the most beautiful complex systems humanity has constructed. We use data about route updates in BGP to model a dynamic system based on the topology of the AS-level graph. We also use metadata about the timestamps of updates to propose an anomaly detection method based on the burstiness of these updates.

The next subsections introduce general research questions and specific contributions in each part and chapters of this dissertation.

1.3.1 Part I: Anomaly Detection in the Society

Beginning with email and Twitter data surrounding specific events, we can ask the following questions:

- **Do important events change communication patterns?**
- **Is the community structure of communication networks robust against external events?**
- **Can we leverage the distribution of links in the community partition for event detection?**

1.3.1.1 Chapter 3: Community-Based Event Detection in Temporal Networks

We propose a method for detecting large events based on the structure of temporal communication networks. Our method is motivated by previous findings that show that patterns of information diffusion with respect to network community structure can reveal the viral nature of contagion. Namely, we hypothesize that large events trigger viral information cascades that cross community boundaries. By comparing the amount of communication within and across communities, we show that it is possible to detect events, even when these events do not trigger significantly larger communication volume. We demonstrate the effectiveness of our method using two examples—the email communication network of Enron and the Twitter communication network during the Boston Marathon bombing.

Figure 1.1 illustrates the key idea of the proposed method. When there is no global event, the communication happens mostly within each community (see Fig. 1.1(A)). However, when an global event occurs, it spreads virally, crossing community boundaries. If the event is relevant to most people, then usual community boundaries shaped based on topics do not constrain the communication anymore.

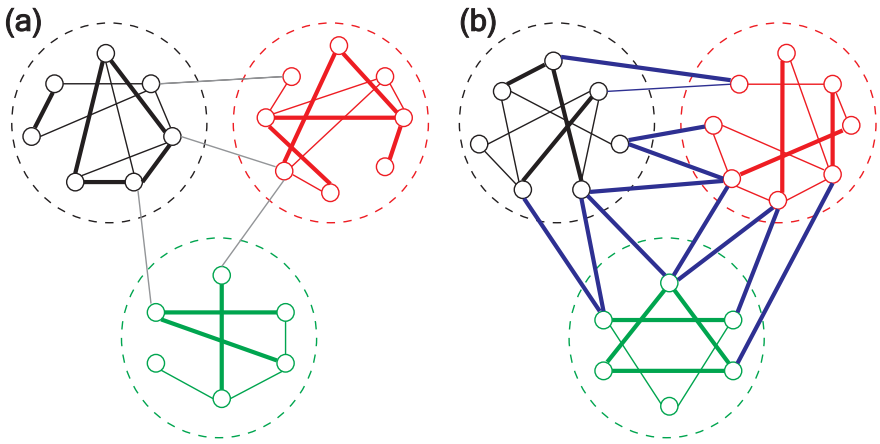

Figure 1.1: Schematic representation of the proposed event detection method. Both networks have the same communities but different patterns of communication within and across them. (A) When there is no global event, most communication occurs within communities. (B) When a global event occurs, more communication may happen across communities because of the global relevance of the event and because of its virality.

The results of this project were initially presented as a conference abstract [Moriano et al., 2017a] and then were published as a journal article [Moriano et al., 2019a] in collaboration with Jorge Finke and Yong-Yeol Ahn. Pablo Moriano is the primary researcher on both works and contibuted all the analysis and figures therein.

1.3.2 Part II: Anomaly Detection in an Organization

Internal events in an organization may influence behaviors in their employees. Research from sociology suggests that the motivation to attack from insiders is associated with internal circulating and often negative information known as precipitating events. Our work on insider threat addresses the following questions:

- **Can we detect the effect of precipitating events in the commit behavior of developers?**
- **What topological properties in the contributions graphs change with the announcement of precipitating events?**
- **Are changes in the commit behavior of developers closely correlated with precipitating events?**
- **Can we leverage the distribution of commits in the community partition for event detection?**

1.3.3 Chapter 4: Protecting Organizational Assets Through Graph Mining

We developed an innovative approach that captures the temporal evolution of user-system interactions to create an unsupervised learning framework to detect high-risk insider behaviors. Our method is based on the analysis of a bipartite graph of user and system interactions. The graph mining method detects increases in potential insider threat events following precipitating events. Specifically, I identify events that have been shown to increase the incidence of insider threats and then leverage a graph-based method to show the increase of detection of insider events at the time of these events. Events such as mass layoffs or restructuring have been shown in the literature

outside of computer science to increase incidents of insider threat activity, and I use that fact to evaluate the graph theoretical approach. We apply our method to a dataset that comprises interactions between engineers and components in a software version control system spanning 22 years and automatically detects statistically significant events. We find that there is statistically significant evidence for increasing anomalies in the committing behavior following precipitating events. Although these findings do not constitute detection of insider threat events per se, they reinforce the idea that insider operations can be motivated by the insiders' environment and detected with the proposed method. We compare our results with algorithms based on volume-dependent statistics showing that our proposed framework outperforms those measures. This graph mining method has potential for early detection of insider threat behavior from user-system interactions, which could enable quicker mitigation.

Figure 1.2 shows the abstraction that we used for characterizing interactions between developers and repositories using bipartite graphs (above). We rely on the developer projection (below) to measure the proportion of commits that permeate different groups of developers. We show that after a precipitating event is announced there is a diversification of commits.

The results of this project were initially published as a conference paper [Moriano et al., 2017b] where the work received a Best Paper award, and an extended version of it was published as a journal article [Moriano et al., 2018a] in collaboration with Jared Pendleton, Steven Rich, and L. Jean Camp. Pablo Moriano is the primary researcher on both works. The initial research question, the analysis, the figures, and the discussion were all contributed by Pablo Moriano.

1.3.4 Part III: Anomaly Detection on the Internet

The Internet, although extremely robust, is notoriously susceptible to attack by means of BGP. Incipient attacks can be detected by analyzing anomalies in the expected behavior of announcements in the BGP protocol. Here we analyze the variation in the generation of BGP anomalies, the topology of the AS-level graphs, and the dynamics of inter-arrival times of updates to address this challenge. We also ask the following questions and make the following contributions:

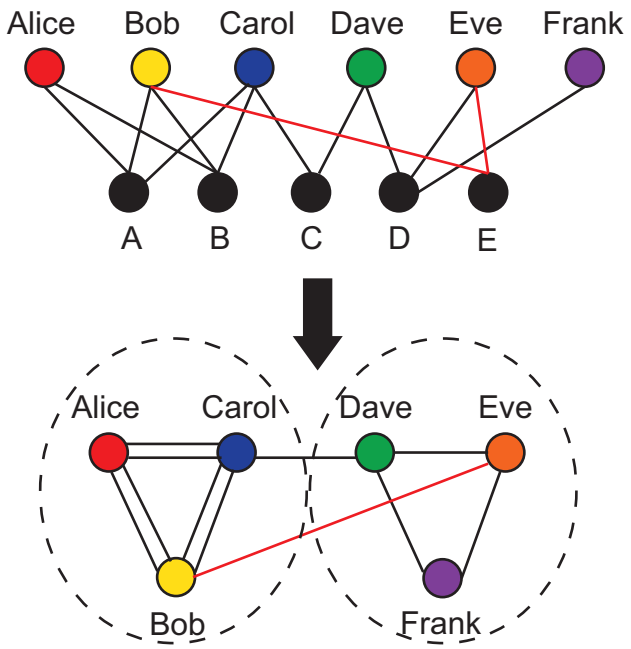

Figure 1.2: Malicious activity in the bipartite graph. The top panel represents an unusual interaction between Bob and Eve with component "E." The bottom panel represents the corresponding one mode projection graph with the anomalous edge crossing communities. Please refer to Figure 1.1 for a more abstract schematic illustrating the method.

- **Can variations in the generation of routing anomalies be explained by macroeconomic indicators?**
- **Are routing anomalies only the result of limited technical capacity?**
- **Are routing anomalies potentially associated with ecrime activities and national intelligence operations?**
- **Is the AS-level graph structure impacted by large-scale routing events?**
- **Are the path lengths in the crust of the AS-level graphs changing during large-scale routing events?**
- **Can we distinguish between large-scale incidents and no incidents by analyzing AS-level graphs?**

- **Do the inter-arrival times of BGP updates change when an large-scale event is incipient?**

- **Can we leverage changes in the inter-arrival times of updates to generate signatures of anticipated detection?**

1.3.5 Chapter 5: Macroeconomic Analysis of Routing Anomalies

We perform a longitudinal study of routing anomalies to investigate the factors that motivate, or at least correlate with, them. We use macroeconomics to examine anomalies as possibly due to incompetence, ecrime, or intelligence operations. Our hypotheses are derived by leveraging three theories from criminology and global measures of technology adoption. We found that exports in technology were not statistically significant, undermining the argument for incompetence. We also found support for the possibility that anomalies are driven by crime, specifically for the guardianship and relative deprivation theories of crime. In addition to these findings from regression analysis, clustering indicates that civil conflict and surveillance are associated with the disproportionate origination of routing anomalies. This supports the possibility of use of routing anomalies for national intelligence.

Figure 1.3 shows the geographic location of the countries that generate a significant number of routing anomalies and poor governmental practices (highlighted in red). Note that countries depicted with higher intensity (darkness) are evaluated by consistent global measures as having lower integrity of governmental institutions. Interestingly, most of these countries have been involved in civil protests and threats of civil war as has been reported in [Dainotti et al., 2011]. In particular, Syria has been (during the observation period) constantly categorized as a country with a high ratio of anomaly/ASes. Other countries in the Middle East that appear on this chart, Iraq and Yemen, also have been embroiled in conflict. The analysis highlights other countries in Africa including Burundi, Comoros, Congo, Guinea, and Liberia, as well as Timor-Leste[3] in Asia. Please note the

[3]Timor-Leste was previously known as East Timor until winning independence from Indonesia and then being established as a sovereign state in 2002.

strategic location of Comoros for undersea cables.

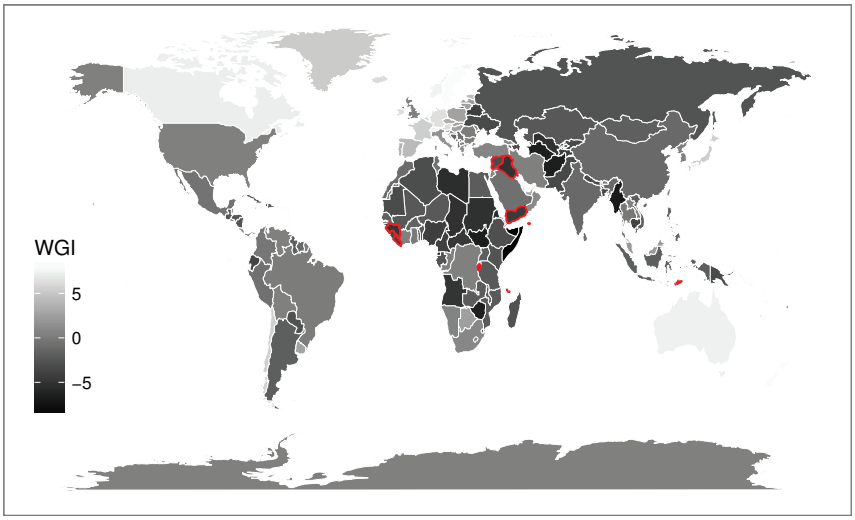

Figure 1.3: Geographic location of the countries with considerable high number of anomaly/ASes and poor governmental practices during the observation period.

The results of this project were initially published as a conference abstract [Moriano et al., 2016] where the work received a Best Poster award. It was then extended into a journal article [Moriano et al., 2017c] in collaboration with Soumya Achar and L. Jean Camp. Pablo Moriano is the primary researcher on both works, defined the method, completed the analysis, and created the figures therein.

1.3.6 Chapter 6: Characterization of Internet Routing Anomalies Through Graph Mining

We explore the idea of reconstructing the AS-level graph for three large-scale routing incidents and evaluate the topological properties of the graphs before, during, and after these events. Using the AS graph topology, we aim to illustrate that the incidents were visible as anomalies before they were widely diffused. The three incidents we examined were the Indosat hijack in April 2014, the Telecom Malaysia leak in June 2015, and the Bharti Airtel Ltd. hijack in November 2015. There were immediate changes in the topological features in the subgraphs, but not the graph as a whole, during the course of these events. Specifically, when the AS-level graph is examined using k-shell decomposition, there are topological changes in the crust in path length and clustering

measurements. The k-shell decomposition distinguishes between the core and periphery graphs. In this k-shell decomposition, the core consists of ASes with at least connectivity k, with the crust consisting of those ASes which have less than k connectivity. Although anomalous behavior was not observable in the core graph, the events are immediately apparent on the crust. However, these changes were not statistically significant. Our hypothesis is that, in graph-theoretical terms, these incidents require the initiators to move closer to the core, away from the periphery, and the concentric impacts of the disturbances are visible as these move across the crust.

Figure 1.4 shows the intuition behind our hypothesis. Under the assumption that the core of the AS-level graphs do not change much and that the attacker is at lower shell levels, malicious updates will tend to change the topology of the graphs. In particular, by analyzing k-crust graphs (generated by aggregating all the previous k-shells), we study whether there are significant changes in the topology of this graphs under the presence of large-scale incidents. We found however that is challenge to capture this behavior based on the busier dynamics of BGP updates (that hinder to find the difference between regular and anomalous).

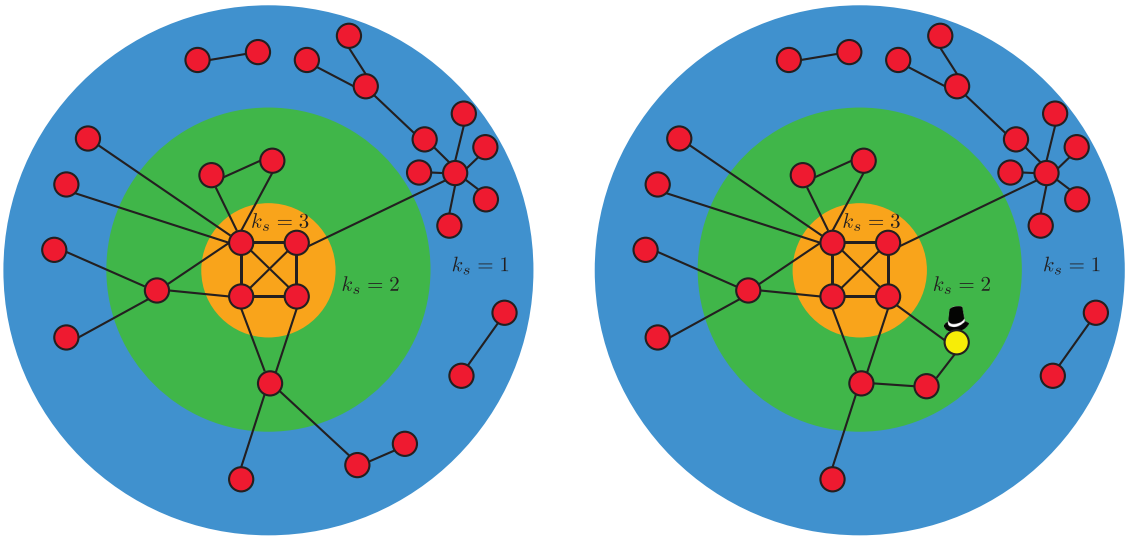

Figure 1.4: A schematic representation of the graph under k-shell decomposition for different values of shell k_s. The AS-level graph under normal conditions is on the left and the same topology when a BGP hijack is committed is on the right. The hijacker is highlighted with the black hat.

The results of this project were initially published as a technical report [Moriano et al., 2017d] then presented as a conference abstract [Moriano et al., 2018b] in collaboration with Raquel Hill

and L. Jean Camp. Pablo Moriano was the primary researcher on both works, implemented the method, completed the analysis, and created the figures therein.

1.3.7 Chapter 7: Bursty Announcements for Early Detection of BGP Routing Anomalies

In this work we propose a method for early detection of large-scale disruptions based on the analysis of bursty BGP announcements. We hypothesize that the occurrence of large-scale disruptions are preceded by bursty announcements. Our method is grounded in analysis of changes in the inter-arrival times of announcements. BGP announcements that are associated with disruptive updates tend to occur in groups of relatively high frequency, followed by periods of infrequent activity. To test our hypothesis, we quantify the burstiness of inter-arrival times around the date and times of three large-scale incidents: the Indosat hijacking event in April 2014, the Telecom Malaysia leak in June 2015, and the Bharti Airtel Ltd. hijack in November 2015. We show that we can detect these events several hours prior to when they were originally detected. We propose an algorithm that leverages the burstiness of disruptive updates to provide early detection of large-scale malicious incidents using local collector data. We describe limitations, open challenges, and how this method can be used for large-scale routing anomaly detection. This detection leverages the fact that for a large-scale event to be implemented effectively, the attacker must send out announcements widely and quickly.

Figure 1.5 shows the the differences in terms of inter-arrival times for two different random processes X_1 and X_2. In X_1, events follow each other at relatively regular time intervals. In contrast, X_2 shows a bursty sequence of events in which there are a burst of events followed for period of relatively long inactivity. We relied on this last observation to hypothesize that for incipient large-scale routing events, the sequence of route updates from ASes involved in the incidents follow a bursty pattern. We then quantify this burstness to propose an event detection algorithm by relaying on measuring the inter-arrival times of route updates.

This manuscript is under revision at SIGCOMM 2019 in collaboration with Raquel Hill and L. Jean Camp. Pablo Moriano was the primary researcher on both works, implemented the method,

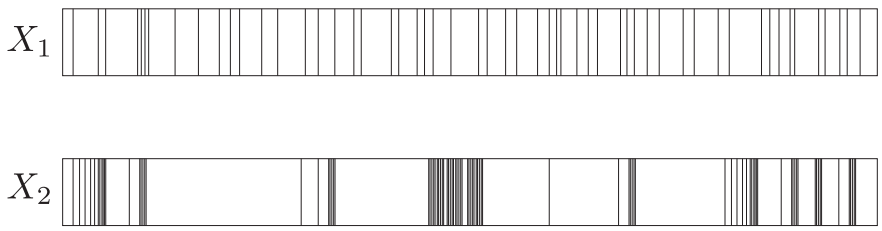

Figure 1.5: Difference between events produced a non-bursty sequence (above) and a bursty one (below).

completed the analysis, and created the figures therein.

1.4 Research Impact

Many computer security challenges today occur at a large-scale. Moreover, security data is intrinsically dynamic meaning that the threat usually rapidly evolves between analysis and the use of countermeasures in detection. The proposed research lays the foundation for extracting meaningful signatures from longitudinal security to detect anomalies using data-driven and network science methods. With insights about the mechanisms that underlie the generation of computer security anomalies, the proposed methods aim to elucidate the creation of countermeasure tools for identification of anomalies in the context of Internet control-plane, email and social media networks, and the insider threat.

1.5 Outline

The remaining chapters of this dissertation are structured as follows. Chapter 2 reviews the relevant literature on Internet control-plane functioning, graph-based anomaly detection, and its intersection with computer security challenges such as the insider threat. In the next five chapters—from Chapters 3 through 7—this dissertation shows the work toward data-intensive hypothesis-driven analysis, including (i) the development of a graph mining method to identify large events in email and social media datasets (in Chapter 3); (ii) an extension of the previous method to take into account the intrinsic interaction between developers and code repositories in a bipartite graph

scenario to identify highly-risky behavior (as the one posed by insiders) (in Chapter 4); (iii) a spatio temporal clustering analysis to understand the country-level behavior of BGP control-plane incidents (in Chapter 5); (iv) the application of k-core decomposition analysis to understand large-scale routing anomalies leveraging AS-level temporal graphs (in Chapter 6); and finally, (v) the development of a method for early detection of large-scale disruptions based on the analysis of bursty BGP announcements (in Chapter 7). Chapter 8 revisits the thesis statement, presents concluding remarks, and suggests areas for future research.

Beyond the specific hypotheses, the nuanced understanding enabled by a temporal graph-based approach also enables an analysis of ground truth. By definition, security data record only those incidents which are detected. The creation of novel graph-based indicators enables the security community to characterize the scope and scale of possible false negatives. The hypothesis-driven temporal graph theory approach here makes it feasible to ground the Internet-scale anomalies in the social and political context, leveraging communities and geography without suffering from over-fitting or spurious correlations while leveraging vast amounts of data.

2 Related Work

"If you cannot explain something in simple terms, you don't understand it well enough."

— Richard P. Feynman

In this chapter, I review previous work related to the topics presented in this dissertation proposal. The related work section has been grouped into sections where I 1) discuss the intersection between anomaly detection and security data; 2) give an overview of graph based anomaly detection with a particular emphasis on event detection in temporal graphs; 3) cover the basis of BGP functioning, including traditional approaches to secure the Internet control-plane and its manipulation for ecrime and political operation purposes; and 4) explain how insider threat has been framed as an anomaly detection problem by characterizing their actors and the use of graph based methods to perform event detection of high risky behavior.

2.1 Security as Anomaly Detection

The ability to discover irregular behavior in networked systems depends on our understanding of how regular mechanisms lead to observable outcomes. Identifying such behavior is usually refereed to as the problem of anomaly detection. Anomaly detection algorithms often prevent the occurrence of critical events or undesired system conditions [Chandola et al., 2009]. Applications range from network intrusion [Ding et al., 2012], credit card fraud [Bolton and Hand, 2001], tax evasion [Bolton and Hand, 2001], malware detection [Invernizzi et al., 2012], and route hijacking [Shi et al., 2012].

In a number of applications, an anomaly is viewed as an "observation that differs so much from other observations as to arouse suspicion that it was generated by a different mechanism" [Hawkins,

1980]. Detection techniques define the regular behavior of a system, against which unusual patterns are evaluated. Characterizing regular behavior is often a prerequisite to identify these anomalies. Consider for example the case of email exchanges in an organization. On the one hand, receiving huge volumes of emails may represent an anomalous event, e.g., a DDoS attack in an attempt to overflow a mailbox. On the other hand, an increase in emails, as a particular deadline approaches, may not represent an anomaly. Variations in the transmission of information during certain dates under particular circumstances may be the norm [Fond et al., 2014a]. Identifying which stationary trends allow us to design reliable detection algorithms remains an open challenge.

Traditional anomaly detection methods deal with unstructured data, i.e., data where samples are not explicitly interrelated. A sample generally represents a multidimensional vector of features or attributes. Most detection methods are based on (i) classification, which trains a classifier from labeled training samples; anomalies are distinguished based on inherited knowledge [Abe et al., 2006, Janssens et al., 2009]; (ii) distance measures, which calculate the similarity between samples; anomalies are samples far from their closest neighbors [Aggarwal and Yu, 2001]; (iii) clustering measures, which verify whether samples belong to emerging centroids; anomalies are samples far from any centroid [Wang and Wilkes, 2012]; (iv) statistical methods which analyzes whether samples occur in high probability regions of a stochastic model; anomalies occur in the low probability regions of the model; (v) information measures, which estimate the information content; anomalies are samples that induce irregularities into that content [Smets and Vreeken, 2011]; and (vi) spectral measures, which find locally sparse lower dimensional projections of the samples; anomalies differ from regular samples on the projected space [Shyu et al., 2006].

2.2 Graph-Based Anomaly Detection

Classical approaches for anomaly detection assume that samples are unstructured. However, in a number of situations, samples are explicitly interrelated. Network representations are a natural way to model the relationships between samples, providing an abstraction that captures the structure of these relationships [Akoglu et al., 2015, Newman, 2003].

In general, the problem of anomaly detection in structured data has been tied to a three-tiered taxonomy [Ranshous et al., 2015]. Algorithms are classified based on (i) the temporal framework; (ii) the type of anomaly; and (iii) the method underlying the approach. Classification based on the temporal framework takes into account two approaches. The first approach focuses on static graphs, that is, on a snapshot of the network at a specific time. The second approach focuses on a sequence of graphs, that is, on the evolution of a network over time. Classification based on the type of anomaly focus on identifying anomalies on the topology (i.e., anomalous nodes, edges, or subgraphs) or on the occurrence of events that affect the topology [Ranshous et al., 2015]. Finally, the classification by the method has a similar sub-classification as for unstructured data, based on the mathematical framework used for detection.

2.2.1 Graph-Based Event Detection

Temporal networks are represented by an ordered set of graphs, called a graph stream. Subsets of this stream are referred to as graph segments [Sun et al., 2007]. There are five general approaches to design event detection algorithms in temporal networks [Ranshous et al., 2015]. First, compression methods use the minimum description length principle to achieve a compact graph representation [Rissanen, 1978]. The idea behind this principle is to find regularities in a data set and to reduce the graph to a compressed representation. An anomalous event is reported based on the difficulty of achieving compressed representations. The goal is to reduce the entropy of the binary representation of the adjacency matrix so as to minimize the cost of encoding [Sun et al., 2007]. An event is detected when the change in the encoding cost of adding a new graph to a graph segment is greater than the cost of the newly aggregated graph. In other words, if the newly aggregated graph inhibits compressibility (because the vertex partition is different from the other graphs in the segment) the algorithm reports the occurrence of an anomalous event. Compression-based methods do not make an assumption or any assumption about the underlying distribution of the data. Their performance is highly dependent on the information theoretic measure, performing well when a significant number of anomalies occur [Chandola et al., 2009]. The average rankings

of detected anomalous events drops notably as the regularity of the normal records increases [Noble and Cook, 2003].

Second, decomposition methods analyze the spectral properties of a matrix or tensor representation of a graph stream. This approach studies regular patterns in selected eigenvectors, eigenvalues, and singular values. For example, the work in [Akoglu and Faloutsos, 2010] analyzes the texting behavior of a mobile network, where nodes represent users and edges represent sent messages. It builds a correlation matrix of the behavior between all pairs of nodes over a fixed time frame, aiming to extract a sequence of features. The algorithm reports an anomalous event based on deviations (i.e., a low similarity) between the principal eigenvector of the current graph and the aggregated graph during the past time frame.

In a closely related work, nodes that share similar structural properties are associated with different so-called structural roles [Rossi et al., 2013]. The degree of membership of a node to a group is estimated through non-negative matrix factorization and maximum description length, which defines the roles of nodes [Henderson et al., 2012]. To detect anomalous events, the algorithm identifies nodes that have unusual structural transition patterns (i.e., mixed membership). In particular, a large number of role changes over time suggests the occurrence of anomalous events. These decomposition-based methods are suitable for handling high dimensional data sets (e.g., when multiple type of interactions are present). They generally perform well when regular and anomalous behavior are separable in lower dimensional spaces.

Third, distance measure methods evaluate distance between graphs as a metric to identify anomalous events. The distance between two graphs is generally calculated either based on the maximum common subgraph or the graph edit distance (GED), i.e., the number of operations needed to convert one graph into another [Shoubridge et al., 2002]. Two consecutive graphs with a significant distance between them raises an alarm. Variations in these algorithms are usually reflected in the function used to calculated the distance. These methods have been used to extract time series representations of the distance between topological features, which are modeled as an autoregressive–moving-average (ARMA) processes, i.e., by autoregression and moving average

factors, used to create a statistical model of the differences. The algorithm detects an anomalous event if the residuals between the expected and empirical output differences are greater than a certain threshold [Pincombe, 2005].

Along these lines, the work in [Koutra et al., 2013] proposes a similarity score to compare pairwise node affinities, i.e., the influence between corresponding nodes of consecutive graphs. Affinities are calculated using a variation of Belief Propagation [Koutra et al., 2011]. Pairwise similarity matrices are compared using a variation of the Euclidean distance. Distance-based methods do not make any assumption regarding the generative distribution of the data. Their performance generally depends on the particular function used to compute the distance measure.

Fourth, statistical methods are based on constructing statistical (parametric or non-parametric) models (e.g., graph likelihood or the distribution of the eigenvalues) to identify deviations from models. Variations in these algorithms usually consider types of models, different techniques to construct the model, different network properties, and different criteria for defining anomalies [Ranshous et al., 2015]. In particular, the work in [Aggarwal et al., 2011] proposes a reservoir sampling method that maintains structural summaries of the graph stream. Anomalies represent edges between subgraphs that rarely occur. Every time a new graph is added to the stream, the algorithm calculates the likelihood of appearance of particular graph objects that contain unusual edges. A single score is associated for each graph added to the sequence, based on the likelihood of individual edges.

Similarly, the work in [Hirose et al., 2009] introduces an algorithm that aims to detect anomalous events in the distribution of the largest eigenvalues of the adjacency matrix of the vertex-to-vertex correlation matrix, i.e., based on local-vertex features. The algorithm identifies an event if the incoming graph generates a correlation matrix whose largest eigenvalue deviates from the expected distribution. Probabilistic methods propose a solution for anomaly detection when assumptions about the underlying data distribution hold true. Computational complexity becomes a limitation when more robust models are used to analyze the data.

The last category comprises community-based methods. These methods propose algorithms

that focus on analyzing the formation of community structures. The idea behind these approaches is to report an anomalous event whenever there is a significant change in any of the communities. For example, the work in [Duan et al., 2009] quantifies the similarity (using the Jaccard coefficient) among partitions (a division into clusters, such that every vertex belongs to one cluster) between the current graph and past graph segments. A similarity below a certain threshold indicates the occurrence of an anomalous event, i.e., the current graph is not similar to the graphs that are present in the existing graph segment. Similarly, a model-based approach for event detection is introduced in [Peel and Clauset, 2015], in which the community structure is captured using a generalized hierarchical random graph model (GHRG). The model infers a dendrogram of the community structure at different scales. To detect anomalous events, the algorithm identifies changes in the parameters of a fitted model using a likelihood ratio test. Community-based methods are flexible with respect to the type of community detection algorithm.

One of the main advantages of community-based methods is that they mitigate topological biases, i.e., are less dependent on the density of links. For example, in an email or Twitter communication network, new users and interactions are constantly being added and sent. Communication traffic tends to increase or drop based on the time of the time of the day or date. Despite this behavior, there is no anomalous event. It might be a regular pattern of the system [Fond et al., 2014a]. In a different scenario, consider a hypothetical network with a predefined community structure, i.e., a network in which every node has been assigned to a specific group that is defined through a partition. If a fixed number of edges is added randomly to this network, this might not necessarily change the community structure. This does not hold for density dependent measures, including, GED, degree distribution, and clustering coefficient [Fond et al., 2014b]. Of all the methods described above, community-based methods are more robust against pinpointing anomalous events that can be identified by measuring the density on links. However, the computational complexity of the community algorithms is often a challenge. An overview of community detection is offered to the reader below.

2.2.2 Community Structure

Communities are groups of densely connected nodes in a graph. The community structure of a graph captures the interaction between nodes belonging to clusters (also known as modules) that are connected internally but not externally. Such clusters might be derived from different types of interactions between nodes, e.g., common interests, homophily, geographical distance among others [Girvan and Newman, 2002, Newman and Park, 2003, Moriano and Finke, 2014]. Figure 2.1 shows a schematic representation of a graph with three communities.

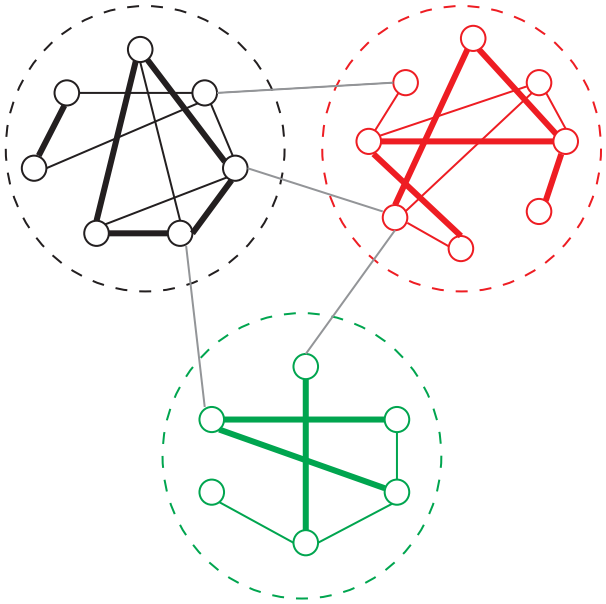

Figure 2.1: A graph with three communities enclosed by the dashed circles.

The problem of detecting communities in graphs is of extremely importance across multiple domains. They include applications that span the analysis of social interactions on the Web [Kleinberg and Lawrence, 2001], the understanding of functional modules in the brain [Bullmore and Sporns, 2009], and in the optimization of routes in large road networks [Song and Wang, 2011]. The challenge of detecting communities, however, is nontrivial [Radicchi et al., 2004]. Different methods for detecting communities have been proposed, including, modularity optimization [Newman, 2006], the Louvain method [Blondel et al., 2008], Infomap [Rosvall and Bergstrom, 2008], clique percolation [Palla et al., 2005], and link clustering [Ahn et al., 2010]. The interested reader

in more details about community detection methods is referred to the extensive review in [Fortunato, 2010]. In this dissertation, the Infomap method was used to compute the community structure of graphs. Details about the Infomap method are discussed below.

Infomap

The Infomap community detection method is based on the notion of codifying the paths of a random walker in a graph. In particular, it assumes that a random walker is more likely to be contained inside communities than to travel between them. The purpose of the method is then to minimize the description length of the random walker's movements in the graph. This notion is captured through the map equation (equation (2.1)) that serves as the objective function in the optimization problem [Rosvall and Bergstrom, 2008].

Specifically, let M be a module partition of n nodes and m modules. It is assumed that each node is assigned to a single module. The map equation $L(M)$ provides an estimate of—given an infinite random walk on a graph partitioned according to M—the average number of bits per step required to describe that walk.

$$L(M) = q_{\curvearrowright} H(\mathcal{Q}) + \sum_{i=1}^{m} p^i_{\circlearrowright} H(\mathcal{P}^i). \tag{2.1}$$

In equation (2.1), $q_{\curvearrowright}$ is the probability that the random walker switches modules per step and $H(\mathcal{Q})$ represents the entropy of modules names. Note that these two terms capture the movement between nodes. In addition, $p^i_{\circlearrowright}$ captures the sum of probabilities of movements inside the module along with the probability of exiting i and $H(\mathcal{P}^i)$ captures the entropy of intra-module movements. Note that the last two terms characterize the movements within the modules.

Overall, the optimization process ensures that the partition with the shortest description length is the one that better describes the community structure of the graph with respect to the dynamics occurring on it [Bohlin et al., 2014].

2.3 BGP

The Internet is composed of over 60,000 computers networks, also know as Autonomosus Systems (ASes) [Huston, 2019]. ASes interconnect globally millions of end systems by means of the Border Gateway Protocol (BGP). BGP was introduced for the first time in 1989 in [Lougheed and Rekhter, 1989] allowing the exchange of reachability information between ASes. The current Internet still relies on BGP to route traffic, although on its fourth version [Rekhter et al., 2006]. Because of the inherent concept of trust derived for the first developments of the Internet, ASes route traffic to certain destinations derived a continuous process of diffusion of routes, i.e., update messages, that do not verify authenticity of claims. This makes BGP a simple to protocol but still unsecured. The operation of the Internet at the BGP-level is also referred as the control-plane.

2.3.1 IP Prefixes

Since the last version of BGP, Classless Inter-Domain Routing (CIDR) IP Prefixes are exchanged between ASes to route global Internet traffic [Fuller and Li, 2006]. In CIDR, the range of addresses that a particular network owns are represented by an IP prefix. An IP prefix has two parts separated by an slash, for example, in the IPv4 block

192.168.100.0/22

the more significant bits represent the network prefix that characterize a network or subnet, i.e., this IP prefix spans 2^{22} IP adresses, from 192.168.100.0 to 192.168.103.255. The least significant bits form a host identifier.

IP address space assignation is an administrative process carried out by the Internet Assigned Numbers Authority (IANA) which is a department of Internet Corporation for Assigned Names and Numbers (ICANN). The IANA distributes the pool of IP addresses between Regional Internet Registries (RIRs) which are responsible for different geographical location. Currently, there are five RIRs worldwide, including, AFRINIC (responsible for the African continent), APNIC (responsible for the Asia/Pacific region), ARIN (responsible for the U.S., Canada, and some regions

in the Caribbean), LACNIC (responsible for South and Central America as well as the remaining portions in the Caribbean), and RIPE NCC (responsible for Europe and the Middle East).

2.3.2 AS Numbers

As in the case of IP address assignation, AS numbers (ASNs) are assigned to ASes for their identification by RIRs. Originally, ASNs were two bytes long [Mitchell, 2013]. However, a new proposal to extend this representation to a four byte representation was presented in [Vohra and Chen, 2012] due to number exhaustion.

2.3.3 Prefix Hijacking

Prefix hijacking is the act of rerouting Internet traffic intended to be delivered to another AS through the propagation of erroneous BGP routes. This is the result of the lack of authentication of BGP update messages. It might be because the results of router misconfigurations [Mahajan et al., 2002] or malicious intent [Ballani et al., 2007]. The AS whose route was hijacked is referred as the victim AS. Different attacks can leverage hijacks. [Zheng et al., 2007] describes the purposes of hijacks as observed in the wild as being to (i) create a black hole in the network, (ii) steal the victim AS's identity, (iii) intercept traffic for eavesdropping, or (iv) to conduct any illegal activity such as sending spam.

2.3.4 Detection of BGP Anomalies

BGP anomaly detection approaches are usually classified based on the type of data that is used for the task. In that respect, there are: (i) control-plane, (ii) data-plane, and (iii) hybrid approaches. Control-plane approaches passively monitor BGP updates or routing tables from a distributed set of BGP monitors [Lad et al., 2006, Khare et al., 2012, Sermpezis et al., 2018a]. These approaches look for inconsistencies in the origin of prefixes announced by ASes or unexpected path changes. In particular, the work in [Lad et al., 2006] proposes a Prefix Hijack Alert System (PHAS). PHAS relies on the idea of finding unique prefixes simultaneously originating from multiple ASes—also

referred as Multiple Origin AS conflicts. Once these conflicts are detected, this method filters false positives using additional information from the network operators, e.g., checking announcements of similar prefixes from different ASes that belong to the same organization. In contrast, the work in [Khare et al., 2012] focuses on correlating suspicious route announcements with past network announcements. This method can detect anomalies that have a huge impact, i.e., announcements that pollute a considerable number of paths. Control-plane methods are usually designed to be implemented as a third-party services such as BGPmon [Toonk]. They have been effective in detecting large-scale events but tend to report a large number of false positives [Zhao et al., 2001, Sermpezis et al., 2018b]. To deal with these shortcomings, the work in [Sermpezis et al., 2018a] proposes ARTEMIS (Automatic and Real-Time dEtection and MItigation System). ARTEMIS is an AS self-operated detection system that exploits local configuration and real-time BGP data from public monitoring services. In contrast with previous control-plane approaches, ARTEMIS provides protection among different types of attacks, including man-in-the-middle traffic manipulation, within a minute of detection delay. All the previously mentioned methods are reactive and notify routing anomalies after the incident occurred. The proposed method belongs to the control-plane category. In contrast with previous methods, it relies on analyzing real-time BGP updates from the route collector perspective and is able to *anticipate* when a large-scale event is going to occur with several hours of anticipation. Our method is able to detect in advance a wide variety of attacks including traffic interception and route leaks.

Data-plane approaches use ping/traceroute to detect anomalies in the route of data [Zheng et al., 2007, Zhang et al., 2010]. These approaches rely on monitoring the reachability of routes from the victim to detect anomalies. The work in [Zheng et al., 2007] proposed a distributed scheme for detecting BGP anomalies based on departures of hop count stability and AS path similarity. Following this methodology, the work in [Zhang et al., 2010] proposed iSPY. iSPY generates an alarm every time the reachability of a predefined prefix is not observable from multiple vantage points. Data-plane approaches are able to pinpoint suspicious path changes in the traffic which results in higher detection accuracy. However, they do not scale well since they require a considerable num-

ber of active measurements for characterizing regular paths and have large latency [Xiang et al., 2011]. Data-plane approaches are complementary to the proposed method, but they are reactive in terms of being able to detect anomalies once they are widely spread and do not allow the ability to anticipate when an event is incipient.

Hybrid approaches have been developed to address the limitations of exclusively control- and data-plane methods [Shi et al., 2012, Hu and Mao, 2007]. The main idea behind hybrid approaches is to use control-plane inconsistencies to inform data-plane measurements, i.e., by exploring the reachability of packets in a particular network. The work in [Hu and Mao, 2007] explored this idea by proposing a framework that launches data-plane probes only when anomalous update messages are received. This system was intended to be used as customized software installed in the routers. Following this idea, the work in [Shi et al., 2012] introduced Argus. Argus is an automated system that detects prefix hijacking and deduces the origin of the anomaly. Argus is based on pervasively correlating control- and data-plane data during a given time period to detect anomalies including sub-prefix hijacks. Following Argus, the work in [Schlamp et al., 2016] introduces HEAP. HEAP relies on the idea of processing update messages to find malicious hijacked prefixes and then scanning the network to find SSL/TLS-enabled hosts. These enabled hosts allow the comparison of public keys prior and during an event, which is the basis for detection of subprefix hijacks.

2.3.5 Mitigation of BGP Anomalies

Several proposals to secure BGP are based on the use of public-key cryptography for the authentication of route announcements [Kent et al., 2000, Ng, 2004, Lepinski and Sriram, 2017]. Cryptographic-signed messages allow the verification of the identity of ASes that claim a certain route. They are based on RPKI for assignment and distribution of public keys [Lepinski and Kent, 2012]. RPKI designates a hierarchy of authorities based on RIRs (regional Internet registries) to allocate and authorize IP space in BGP through the use of digital signatures and public key certificates. RPKI allows secure origin authentication.

The use of RPKI alone does not require changes in the BGP protocol. RPKI is an out-of-band

mechanism in which routers download information for decision making and does not require the use of online cryptography. However, there are reasons that limit the scope of RPKI for securing BGP. Researchers have debated the agreement of a trusted Certificate Authority [Cooper et al., 2013], difficulties to correctly configure RPKI [Wählisch et al., 2012], a general lack of commitment and incentives to lead their implementation [Gill et al., 2011], and its inefficacy in the face of certain types of attacks, e.g., path shortening attacks [Goldberg, 2014].

To remedy limitations with respect to the type of attacks that can be undetectable with only origin validation, path validation proposals have been proposed. As opposed to origin validation, path validation proposals authenticate every AS in the path of a corresponding announcement. The work in [Kent et al., 2000] proposed secure BGP (S-BGP) to validate path attributes in BGP updates messages. Information in S-BGP is validated in RPKI. This is done through the use of attestations, i.e., signed messages that verify the authenticity of route announcements. Address attestations are statements signed by known authorities that map Autonomous System Numbers (ASNs) to prefixes to verify that the ASes originating the route were eligible to do so. Route attestations are statements signed by ASes and operationalized in the AS-PATH attribute. They are used for each AS in the path attribute to confirm that the next AS in the path has received the announcement and was the right to forward it. S-BGP provides full authentication of origin and path through attestations. Along the same lines, another proposal is secure origin BGP (soBGP) [Ng, 2004]. soBGP relies on RPKI for handling public keys to soBGP speaking routers, maintaining certificates of routing policies, and authentication of IP space and ASes. soBGP relies on a graph topology database to validate policy interactions between ASes. Update messages violating AS topology policies are dropped. Note that the graph topology database used in soBGP is relatively static given that the topology will only change when there is a change in the policy agreements. In contrast, S-BGP performs attestations every time there is a new update.

Another proposed standard is BGPSec [Lepinski and Sriram, 2017]. BGPSec builds on RPKI to distribute and manage cryptographic keys that are used to sign and authenticate every AS on the path of a corresponding announcement. In contrast to S-BGP and soBGP, BGPSec is an in-

tegral part of the BGP protocol requiring online cryptography in which routers sign and verify every message that they sent. This creates computational overheads and requires routers hardware upgrades. This economic incentive, for network operators, makes it difficult to think of in its fully implementation [Goldberg, 2014]. In the context of partial deployment, the work in [Lychev et al., 2013] shows that BGPSec alone does not offer a significant security improvement when compared with only RPKI usage under certain routing policy scenarios.

Although RPKI and path validation proposals are able to protect BGP against a wide variety of attacks, they fail when trying to avoid the adoption of leaked routes. A route leak occurs when an AS announces valid routes beyond their intended scope, i.e., the AS announces a route that is in violation of the receiver, the sender and/or one of the ASes along the preceding AS path [Sriram et al., 2016]. These types of anomalies still can generate blackholes and are even prone to traffic interception [Goldberg et al., 2010]. Our proposed method is lightweight and able to detect in advance malicious route manipulation and route leaks.

For a comprehensive surveys on BGP anomaly detection and mitigation methods, we refer the reader to the works in [Butler et al., 2010, Al-Musawi et al., 2017, Mitseva et al., 2018].

2.3.6 Macroeconomics of Security

Much of the research on ecrime has been strictly empirical. For example, Kanich et al. examined the value chain in pharmaceutical spam, where the payment services processors were identified as the weakest link [Kanich et al., 2008]. This technical analysis enabled an effective policy response, significantly decreasing the prevalence of pharmaceutical spam. Christin focused on monetary returns rather than the mechanics of the attacks [Christin, 2013]. That analysis of underground markets illustrated the stability of these markets using the example of online narcotics and estimated the revenue of the market operators.

Nation and region of origin have been found to be relevant in analyses of spam, both by industry researchers [Microsoft, 2011] and academics [Garg et al., 2013]. Macroeconomic analysis of malware illustrates the significance of state-level variables, including the economic variables that

we use in the proposed research [van Eeten and Bauer, 2008, Garg and Camp, 2013]. Our analysis also draws on three different criminology theories that have been correlated with the production of spam using macroeconomic techniques [Garg et al., 2013].

Afroz et al. provided a window into the geographic and political distribution of online crime by analyzing linguistic groupings in online criminal marketplaces [Afroz et al., 2013]. Moore et al. examined temporal patterns and value chains in phishing and theft, providing guidance for coordinated law enforcement responses [Moore et al., 2009]. Later work by van Eeten et al. considered the role of industry and national policies in the spam ecosystem [van Eeten et al., 2010]. van Eeten et al. also found adoption of the guidance in Moore et al. to be correlated with a decrease in infected machines for ISPs in Europe.

Other related research investigates the relationship between short-lived BGP hijacks and malicious activities such as spamming and denial of service (DoS) attacks. Specifically, Clayton described pollution of registry databases by attackers [Clayton, 2015]. This attack enabled miscreants to claim official ownership of IP space in order to subvert automatic filtering of invalid route announcements. These misidentified IP addresses were then used in spam and malware campaigns.

Additional empirical analysis had shown that a significant amount of spam is received from IP addresses that correspond to short-lived, possibly hijacked IP prefixes [Ramachandran and Feamster, 2006]. Hu and Mao also discussed the possible use of BGP hijacks for conducting DoS attacks [Hu and Mao, 2007]. The analysis of these types of attacks provides an interesting indirect link among the causal factors that are related to the origin of hijacks. More importantly, these findings suggested that BGP hijacks may be becoming part of the commoditized ecrime ecosystem.

2.3.7 Political Operations

There is evidence of the manipulation of the Internet control-plane with political intent. Earlier work analyzed the technical signatures around particular Internet censorship events, which can be considered political attacks on the Internet. In particular, Dainotti et al. presented evidence on how Internet communications were disrupted—through BGP-based disconnection—in several

North African countries in response to civilian protests and threats of insurrection [Dainotti et al., 2011].

Similarly, the use of deep packet inspection (DPI) has been examined as a political and technical phenomenon across different ASes. In particular, Asghari et al. identified the economic and political drivers that have a significant correlation with DPI [Asghari et al., 2012]. This work found that high levels of governmental censorship and weak privacy protections are associated with a pervasive use of DPI.

Hiran et al. analyzed the impact of the China Telecom incident in terms of the origin of prefixes that were announced [Hiran et al., 2013]. In this incident, China Telecom claimed a significant percentage of the the IPv4 address space [Toonk, 2010]. Their results suggest that most of the hijacked prefixes correspond to organizations in the U.S., followed by organizations in China, South Korea, Australia, and Mexico. After detailed inspection, they found that the U.S. shows a lower rate of hijacked prefixes based on the global distribution of Internet prefixes across the globe. Conversely, countries in the Asia-Pacific region (e.g., China, South Korea, Australia) were more affected by the incident based on the global distribution of prefixes.

Arnbak and Goldberg reported on the manipulation of BGP and DNS protocols to divert U.S. traffic abroad—where it can be collected under a different and more permissive jurisdiction [Arnbak and Goldberg, 2015]. In addition, the manipulation of the BGP protocol to route traffic through unusual paths which result in crossing new and different jurisdictions was also studied in [Benton and Camp, 2016]. For example, Cowie described how traffic intended to be delivered from one Denver ISP to another Denver ISP was subject to a detour (through Iceland) and possibly interception or manipulation [Dyn Guest Blogs, 2013]. It is certainly difficult to believe such a detour was optimal when the endpoints were located in the same city.

These cases identify possible causal factors relating to political intentions in traffic redirection. More importantly, this previous research suggests that BGP hijacks may be becoming an attack vector for intelligence operations.

2.4 Insider Threats as Anomalies

The analysis of the insider threat using temporal graph mining is informed by past research in the characterization of insider threats and detection of insider threat using graph-based approaches. Here, we provide an overview of related works in these two areas.

2.4.1 Characterization of Insider Threats

Much of the research on insider threats have been on the characterization of insiders. In general, two different categorizations have been proposed to classify insiders. The first one focuses on the intention of the attack [Cappelli et al., 2012]. Under this categorization, insiders are classified as (i) malicious, where the insider intentionally causes a negative impact on the confidentiality, integrity, and availability of the information system; and (ii) non-malicious (accidental), where an insider, through action or inaction but no malicious intent, causes harm.

The second categorization is given with respect to the purpose of the attack [Bishop et al., 2014]. With that definition in mind, two types of attacks are defined more precisely, including (i) a sabotage attack in which the insider is able to change the value of an artifact used in the computation of a process; and (ii) a data exfiltration attack in which the insider provides access to artifacts for entities that are not entitled to that access.

In addition to the previous two-tiered categorization, Nurse et al. proposed a unifying framework to characterize insiders based on the motivation behind malicious threats and the human factors related to the unintentional cases [Nurse et al., 2014]. This framework is of particular importance not only because it leverages previous insider threat case studies, but also due to its analysis of behaviors that may lead to attacks and the types of attacks that may be executed. The factors that are proposed to this end encompass precipitating events and the motivation to attack. Similarly the works in [Liu et al., 2008, 2009] provide support for the understanding of malicious insiders' motives and their decision-making process using game theory.

2.4.2 Insider Threat Detection Using Graph-Based Methods

Graph mining techniques have also been used as a tool to understand and identify malicious actions by insiders. Eberle et al. proposed an approach to detect anomalous subgraphs with respect to the number of transformations that a subgraph will need in order to be a reference—the normative or best—subgraph [Eberle et al., 2010]. The approach relies on MDL to quantify the number of required transformations as a criterion of decision [Noble and Cook, 2003]. The authors validated their approach using empirical data on a passport processing scenario. In particular, they were able to identify some bypassable steps in the process of getting a passport, which represents an anomalous structure of unseen edges.

To address the dynamic nature of empirical data, in a recent work, Eberle et al. introduced a method for pattern learning and anomaly detection in streams using parallel processing [Eberle and Holder, 2015]. This work offers a considerable improvement on speedup compared to the previous approach by allowing the processing of dynamic data. The authors validate their approach on empirical data of embassy employee activity in which the threat was information leakage by employees.

Kent et al. used the notion of bipartite graphs—by capturing interactions through authentication logs between users and computers—for assessing network authentication trust risk and cyber attack mitigation [Kent et al., 2015]. In particular, they examined the number of connected components (*i.e.* a subgraph in which any two nodes are connected to each other by a path) in the bipartite graph to assess potential risk of credential stealing and compromise within an enterprise network. They found that the increase in the number of connected components in the bipartite is associated with a reduction in the risk associated with credential theft and subsequent credential hopping within the network.

Of similar nature, Chen et al. proposed an unsupervised learning model based on social network analysis for detecting anomalous access in collaborative information systems [Chen et al., 2012]. Their approach relied on the quantification of pairwise similarities of nodes in a graph based on their interactions with particular subjects when interactions are made between users and

subjects in a bipartite graph setting. The authors validated their results with patient record access data and Wikipedia edit logs.

3 Community-Based Event Detection in Temporal Networks[1]

"The more original a discovery, the more obvious it seems afterwards."

— Arthur Koestler

3.1 Introduction

In this chapter, we propose a method for detection of large events in temporal communication networks. Our method is based on the the idea that large events initiate information cascades that spread through the inter communication links of the underlying community structure. Our study is based on the email communication network of Enron and the Twitter communication network during the Boston Marathon bombing.

3.2 Problem

Event detection is of crucial importance in many socio-technical networks [Hawkins, 1980]. Events often represent the occurrence of anomalies or undesired outcomes [Chandola et al., 2009] in applications that range from network intrusion [Ding et al., 2012], credit card fraud [Bolton and Hand, 2001], disease outbreak [Wong et al., 2003], fault detection [Basu and Meckesheimer, 2007], and criminal analysis [Lin and Brown, 2003]. However, little attention has been given to the challenge of how to differentiate the effects of anomalies when the "normal" status of the system allows for high variations.

[1]The content of this chapter was initially presented as a conference abstract [Moriano et al., 2017a] and then published as a journal article [Moriano et al., 2019a] in collaboration with Jorge Finke and Yong-Yeol Ahn. Pablo Moriano is the primary researcher on both works and made all the analysis and figures therein.

Consider the case of email exchanges in an organization. An unusual volume of incoming emails does not necessarily mean an attack to the organization's email system. For instance, an increase in the volume of emails, may reflect regular behavior if there is an important deadline. In other words, even large variations in the communication patterns may be normal under particular circumstances [Fond et al., 2014a]. This chapter focuses on identifying topological features that enable the design of reliable detection algorithms.

Traditional anomaly detection methods use multidimensional feature vectors [Abe et al., 2006, Janssens et al., 2009, Aggarwal and Yu, 2001, Wang and Wilkes, 2012, Smets and Vreeken, 2011, Shyu et al., 2006]. However, these methods do not account for the structure in graph data, which can significantly impact the detection task [Akoglu and Faloutsos, 2010, Rossi et al., 2013, Pincombe, 2005]. Here, we propose a new method for detecting events by examining communication patterns based on the theory of information diffusion.

Our method is closely related to existing methods that monitor community structures. For example, the work in [Duan et al., 2009] uses the Jaccard coefficient to quantify the similarity among non-overlapping communities (i.e., network partitions such that every vertex belongs to only one community) between the current and past networks. A similarity value below a given threshold indicates the occurrence of an event, that is, that the current community structure of the network differs significantly from past structures. Similarly, a model-based approach for event detection is introduced in [Peel and Clauset, 2015], in which a hierarchical community structure of the graph is captured using a generalized hierarchical random model. To detect anomalies, the algorithm identifies significant changes in the parameters of a fitted model using a likelihood ratio test.

An advantage of community-based detection methods is their robustness to small topological changes, specifically to fluctuations in the link density [Karrer et al., 2008, Yang et al., 2015]. For example, in email or Twitter networks, users constantly establish new links. Communication traffic tends to vary based on particular times of the day and dates. Such variations represent a regular pattern of the network and should not be consider an anomaly [Fond et al., 2014a].

Our method focuses on communications patterns with respect to community structure. Instead of monitoring changes in the community structure itself, we suggest that the difference between the ratio of inter-community communication and the intra-community communication captures the normal dynamics of information spreading in a network. Deviations from the expected ratio indicate the occurrence of large events [Weng et al., 2013]. The proposed method is less computationally expensive compared to other community-based methods because it does not require computing the similarity between communities of two networks each time. That is, it does not need to calculate the similarity between the community of the current graph and the community that characterizes the growing segment as a detection signature. This makes the proposed method independent of similarity metrics issues and overheads. The proposed method is also independent of generative models that capture the community structure embedded in the data.

3.3 Methods

3.3.1 Data

We use two datasets: the email communication network from Enron and the mention and retweet network from Twitter.

3.3.1.1 Enron Email Communication Network

Enron was of the largest U.S. businesses in the late 90s [Diesner et al., 2005] when it filed for bankruptcy in 2001. The company omitted negative balances and reported inflated profits by allocating losses into fraudulent special purpose entities. The Federal Regulatory Commission published a corpus of Enron's corporate emails [Wilson and Banzhaf, 2009], consisting of 125 153 emails sent by 184 employees[2]. The data is represented by a directed network in which a node is an employee and a link is an email between two employees. The network describes patterns of interaction from January 1999 to June 2002.

[2]Available at `http://cis.jhu.edu/~parky/Enron/`

3.3.1.2 Twitter Interaction Networks During the Boston Marathon Bombing

On April 15th 2013, deadly explosions took place during the Boston Marathon [Sutton et al., 2014]. One of two suspects was shot dead on April 18th and the other suspect was captured on April 19th [Starbird et al., 2014]. We use over 456 million English tweets, posted from April 1st to April 30th, to create a mention and a retweet network.

3.3.2 Network Representation

Consider the sequence of n equal-sized intervals $A = \{A_1, A_2, \ldots, A_n\} = \{A_k\}_{k=1}^{n}$. Let $\mathcal{H} = \{1, 2, \ldots, N\}$ be the set of nodes (e.g., the set of Enron employees or Twitter users). Let $\mathcal{V}(k) \subseteq \mathcal{H}$ be the subset of nodes that interact during interval $A_k = [a_k, a'_k)$. Let $\mathcal{W}(k) = \{\omega_{ij}(k) : i, j \in \mathcal{V}(k)\}$ be a weighted adjacency matrix in which $\omega_{ij}(k)$ captures the number of interactions between nodes i and j. Let $\mathcal{G}(k) = (\mathcal{V}(k), \mathcal{W}(k))$ represent a directed network that takes account of all interactions within interval A_k. The networks comprise reciprocal communications within the largest connected component. Note that $\mathcal{G}(k)$ neglects interval dynamics. Finally, let $G = \{\mathcal{G}(k)\}_{k=1}^{n}$ denote the sequence of the temporal graph slices.

3.3.2.1 Detection Problem

The series G captures the dynamics of the network across time and defines the basis for detection. Let m ($1 \leqslant m < n$) represents the resolution of detection in terms of the number of intervals $A_k, k \in \{1, 2, \ldots, n\}$. For instance, if $m = 2$, then the detection problem is concerned with identifying whether an event occurs within the detection interval $(a_{k-m+1}, a'_k] = (a_{k-1}, a'_k]$, $k = 1, 2, \ldots n$. Let $\bar{n} = \lfloor \frac{n}{m} \rfloor$ be the number of times an algorithm (with resolution m) assesses detection. Let $E \subseteq \{1, 2, \ldots, \bar{n}\}$ represent the intervals at which at least one event occurs (based on the ground truth information). Define $e \in E$ as the index of a detection interval containing an event. Let $\hat{E} \subseteq \{1, 2, \ldots, \bar{n}\}$ represent the set of intervals at which the occurrence of at least one event is reported by the detection method. Similarly $\hat{e} \in \hat{E}$ represents the index at which an event

is reported. The detection problem is defined as follows: Given a series of networks $G = \{\mathcal{G}(k)\}_{k=1}^{n}$ and a detection resolution m, identify the set of intervals $\hat{E}$ containing at least one event.

3.3.3 Algorithm Evaluation

The performance of the algorithm is measured based on identifying the intervals containing events. Specifically, performance is based on the set of time intervals $\hat{E}$ in which an event is reported and the set of intervals E in which events occur (ground truth).

Let $\hat{O}$ be an indicator vector that characterizes the intervals containing an event as per the detection algorithm, i.e., the distribution of events over the set of the $\bar{n}$ detection intervals. Similarly, let O be a vector that indicates the intervals containing an event (or the ground truth). We compare the performance of detection based on true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) rates. In particular, $TP = O \cdot \hat{O}$, $FP = O' \cdot \hat{O}$, $FN = O \cdot \hat{O}'$, and $TN = O' \cdot \hat{O}'$ where "$\cdot$" represents the dot product between vectors, and O' and $\hat{O}'$ represent the complement of O and $\hat{O}$. The detailed pseudo-code for the performance measure is presented in Algorithm 1, which is used to calculate the performance measures of accuracy, precision, recall, and F1 score.

The measure of accuracy is defined as $\frac{TP+TN}{TP+TN+FP+FN}$ and quantifies the proportion of accurately reported positive and negative events (i.e., time intervals classified as positives or not that were correctly classified).

Precision is defined as $\frac{TP}{TP+FP}$ and represents the proportion of positive predictions that have been correctly classified. This means that if a considerable number of intervals are erroneously classified, then detection has low precision.

Recall defines the proportion of actual intervals that have been predicted as positive, defined by $\frac{TP}{TP+FN}$. If an insignificant number of time intervals are classified containing events but they do not, then the algorithm has low recall.

Finally, the F1 score is defined as $2\frac{\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}$ and conveys a balance between precision and recall.

Algorithm 1 Performance ($\hat{E}$, E, $\bar{n}$)

$\hat{O} \leftarrow zeros(\bar{n})$
for $\hat{e} \in \hat{E}$ **do**
  $\hat{O}_{\hat{e}} \leftarrow \{\}$
  for $t \in \{1, 2, \ldots, \bar{n}\}$ **do**
  end for
  $\hat{O} \leftarrow \hat{O}$ OR $\hat{O}_{\hat{e}}$ (element wise)
end for
$O \leftarrow zeros(\bar{n})$
for $e \in E$ **do**
  $O_e \leftarrow \{\}$
  for $t \in \{1, 2, \ldots, \bar{n}\}$ **do**
    $O_e \leftarrow O_e \cup \mathbb{1}_{B_t}(e)$
  end for
  $O \leftarrow O$ OR O_e (element wise)
end for
$O' \leftarrow$ NOT (O)
$\hat{O}' \leftarrow$ NOT $(\hat{O})$
$TP \leftarrow O \cdot \hat{O}$
$FP \leftarrow O' \cdot \hat{O}$
$FN \leftarrow O \cdot \hat{O}'$
$TN \leftarrow O' \cdot \hat{O}'$
return (TP, FP, FN, TN)

Next, we introduce a detection criterion based on the dynamics of the interaction across and within communities.

3.3.4 The Proposed Detection Algorithm

The proposed algorithm aims to define detection signatures based on deviations from the regular process of interactions. In particular, we explore whether variations in the proportion of links across communities are indicators of events. To evaluate the evolution of community structures, the detection algorithm constructs graph segments based on the detection resolutions.

Let the cumulative graph segment of length m be defined as

$$\mathcal{G}_m(k) = (\mathcal{V}_m(k), \mathcal{W}_m(k)) = \bigoplus_{k'=k-m+1}^{k} \mathcal{G}(k') = \mathcal{G}(k-m+1) \oplus \mathcal{G}(k-m+2) \oplus \cdots \oplus \mathcal{G}(k)$$

where $\mathcal{V}_m(k) = \bigcup_{k'=k-m+1}^{k} \mathcal{V}(k')$ and $\mathcal{W}_m(k) = \sum_{k'=k-m+1}^{k} \mathcal{W}(k')$.

Let $T = \{m, 2m, \ldots, \bar{n}m\}$ capture the time intervals at which detection must be assesed. Note that for $k \in T$, the series $\{\mathcal{G}_m(k)\}$ forms a set of non-overlapping cumulative graph segments. Now,

consider an initial graph segment of length m_0, defined by

$$\begin{aligned} \mathcal{G}_{m_0} &= (\mathcal{V}_{m_0}, \mathcal{W}_{m_0}) \\ &= \bigoplus_{k'=1}^{m_0} \mathcal{G}(k') = \mathcal{G}(1) \oplus \cdots \oplus \mathcal{G}(m_0) \end{aligned} \tag{3.1}$$

where $\mathcal{V}_{m_0} = \bigcup_{k'=1}^{m_0} \mathcal{V}(k')$ and $\mathcal{W}_{m_0} = \sum_{k'=1}^{m_0} \mathcal{W}(k')$.

Consider the following assumption:

(A1) The initial graph segment $\mathcal{G}_{m_0}$ can be divided in non-overlapping communities, i.e., the set of nodes that can be grouped into subsets such that nodes belonging to the same subset are densely interconnected [Newman, 2004].

The proposed algorithm reports events based on the proportion of inter- and intra-community links of the graph $\mathcal{G}_m(k)$ with respect to $\mathcal{G}_{m_0}$. For the Enron network, the community partition used as a reference corresponds to a period of $m_0 = 91$ weeks. For the Twitter networks, m_0 corresponds to seven days of user interactions.

To define the set $\hat{E}$, let $C(\mathcal{G}_{m_0}) = \{0, 1, \ldots, c\}$ be a set of unique community identifiers, where $c + 1$ is the number of communities in $\mathcal{G}_{m_0}$. The community to which node $i \in \mathcal{V}_m(k) \cap \mathcal{V}_{m_0}$ belongs (based on $\mathcal{G}_{m_0}$) is given by $c_i : i \to C(\mathcal{G}_{m_0})$. We compute the community partition of $\mathcal{G}_{m_0}$using the Infomap algorithm [Rosvall and Bergstrom, 2008]. Following similar ideas as in [Weng et al., 2013], let $I_{\frown}(\mathcal{G}_m(k)) = \{(i,j) : \omega_{ij}(k) > 0 \wedge (c_i \cap c_j) = \varnothing\}$ represent the set on inter-community links and $I_{\circlearrowleft}(\mathcal{G}_m(k)) = \{(i,j) : \omega_{ij}(k) > 0 \wedge (c_i \cap c_j) \neq \varnothing\}$ the set of intra-community links. Define the inter- and intra-community link ratios as

$$c^m_{\frown}(k) = \frac{|I_{\frown}(\mathcal{G}_m(k)|}{|I_{\frown}(\mathcal{G}_m(k))| + |I_{\circlearrowleft}(\mathcal{G}_m(k))|} \tag{3.2}$$

$$c^m_{\circlearrowleft}(k) = \frac{|I_{\circlearrowleft}(\mathcal{G}_m(k)|}{|I_{\frown}(\mathcal{G}_m(k))| + |I_{\circlearrowleft}(\mathcal{G}_m(k))|} \tag{3.3}$$

Detection focuses on identifying the intervals k, for which $c^m_{\frown}(k) - c^m_{\circlearrowleft}(k)$ exceeds a threshold that is a function of the mean and the standard deviation. We use the sample mean (over the entire period of the study) as the mean estimator because observations seem to resemble a normal

distribution—since hypothesis testing demonstrates that the normal distribution is a good candidate to model the generation of the empirical observations. Moreover, we use the sample standard deviation as the estimator of the standard deviation. The pseudo-code for the detection algorithm is shown in Algorithm 2. The parameter δ controls how many standard deviations are considered to report an interval with an event.

Algorithm 2 Event-Detection (G, m_0, m, δ)

Compute community partition of $\mathcal{G}_{m_0}$
$Y \leftarrow \{\}$ ▷ Array of intra-inter community ratio samples
for k in $\{m_0 + m, m_0 + 2m, \ldots, \bar{n}m\}$ **do**
    Build $\mathcal{G}_m(k) = \bigoplus_{k'=k-m+1}^{k} \mathcal{G}(k')$
    Compute $I_{\curvearrowright}(\mathcal{G}_m(k))$
    Compute $I_{\circlearrowleft}(\mathcal{G}_m(k))$
    Calculate $c^m_{\curvearrowright}(k)$ and $c^m_{\circlearrowleft}(k)$ using eqs. 3.2 and 3.3
    $Y \leftarrow Y \cup \{c^m_{\curvearrowright}(k) - c^m_{\circlearrowleft}(k)\}$
end for
$\mu \leftarrow \frac{1}{\bar{n}} \sum_{y_i \in Y} y_i$
$\sigma \leftarrow \sqrt{\frac{\sum_{y_i \in Y} (y_i - \mu)^2}{\bar{n}-1}}$
$\hat{E} \leftarrow \{\}$
for k in $\{m_0 + m, m_0 + 2m, \ldots, \bar{n}m\}$ **do**
    if $Y(k) >= (\mu + \delta\sigma)$ **then**
        $\hat{E} \leftarrow \hat{E} \cup \{k\}$
    end if
end for
return $\hat{E}$

To measure performance, we compare the measure of $c^m_{\curvearrowright}(k) - c^m_{\circlearrowleft}(k)$ with the respective topological measure derived from $\mathcal{G}_m(k)$, e.g., the volume of interactions—number of links of the cumulative graph segment.

3.3.5 Performance Benchmark

Next, consider the performance of a random algorithm. The output of the random algorithm is given by

$$\hat{R} = (\hat{R}_1, \ldots, \hat{R}_{\bar{n}}) \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(0.5) \tag{3.4}$$

As for the proposed algorithm, the performance of the random algorithm $\hat{R}$ takes into account accuracy, precision, recall, and the F1 score.

3.4 Results

3.4.1 Enron

Using the email data from Enron, we test our method in comparison with other baselines. Figure 3.1 reports the results measured on the email time series between 2000-09-30 and 2002-04-30. First, we evaluate whether the volume of emails correlates with the events associated to Enron's collapse (depicted by the black dashed vertical lines and numbers). These events are described in Table 3.1 and have been identified in [Marks, 2010, Darst et al., 2016]. Figure 3.1(a) shows the weekly volume of emails. The horizontal solid line represents the moving average of emails during the observation period using a window length equivalent to a year of data (52 weeks). Each horizontal red band represents one moving standard deviation from the moving average using the same window length (more intense bands indicate observations that are further away from the mean, based on Algorithm 2). Note that events 1, 4, 5 and 6 lie more than one standard deviation away from the moving average and their occurrence coincides with a burst of emails. However, this relationship does not hold for events 2, 3 and 7.

Event ID	Date	Description
1	2001-05-17	Schwarzenegger, Lay, Milken meeting.
2	2001-07-12	Quarterly conference call.
3	2001-08-03	Skilling makes a bullish speech on Enron Energy Services. That afternoon, he lays off 300 employees.
4	2001-10-16	Enron reports a 618 million third-quarter loss and declares a 1.01 billion non-recurring charge against its balance sheet, partly related to "structured finance" operations run by chief financial officer Andrew Fastow. In the analyst conference call that day, Lay also announces a 1.2 billion cut in shareholder equity.
5	2001-12-02	Enron files for Chapter 11 bankruptcy protection, at the time the largest bankruptcy in U.S. history.
6	2002-02-14	Sherron Watkins, the Enron whistleblower, testifies before a Congressional panel against Skilling and Lay.
7	2002-04-09	David Duncan, Arthur Andersen's former top Enron auditor, pleads guilty to obstruction.

Table 3.1: Enron's event description.

We measure the difference between inter- and intra-community link ratios detailed in Eqs. (3.2)-(3.3). For the Enron dataset, the community partition results from a period of $m_0 = 91$ weeks. Fig. 3.1(b) shows that the occurrence of events, 1 through 7, coincide with peaks in the proposed measure. Even events that do not occur during periods of elevated volume of emails are associated with increased inter-community communication. This result supports our hypothesis that there is a considerable transmission of information through inter-community links when events take place. Note that the activity signal may occur before or after the events. It is natural to expect heightened activity before the "event" in many cases. The "events" in the Enron's case are the public release of certain information. Therefore, it is reasonable to assume that, in some cases, such information had been circulating internally, preceding the actual "event."

Figure 3.1(c) shows the volume classified into six topics [Berry and Browne, 2010]. One topic represents contents associated to daily activities; the remaining ones are associated to Enron's bankruptcy. Most emails are classified into day-to-day activities. For the categories not related to Enron's bankruptcy, there is no association between events and topics, suggesting that the volume of emails does not help us to characterize a detection pattern.

Figure 3.1(d) shows the difference between inter- and intra-community link ratios distinguished by topics. Note the association between events and topics. In particular, emails about day-to-day activities have a similar inter- and intra-community diffusion pattern during the entire observation period, depicted by the flat curve. For topics related to "utility companies difficulties," "Federal Energy Regulatory Commission" and "wire stories about Enron's demise," there is a positive association. In other words, topics that are sensitive to bankruptcy diffuse across communities. The co-occurrence of peaks and events in Fig. 3.1(b) shows that the proposed criterion can be used as a signature for detection.

Figures 3.2 and 3.3 show the performance of detecting events using the proposed criterion against the volume of emails (for different detection resolutions). The detection resolution, denoted by m, describes the number of weeks within which we compute the output of the detection algorithm. We compare detection algorithms using the ROC [Fawcett, 2006] and the PRC [Gordon and

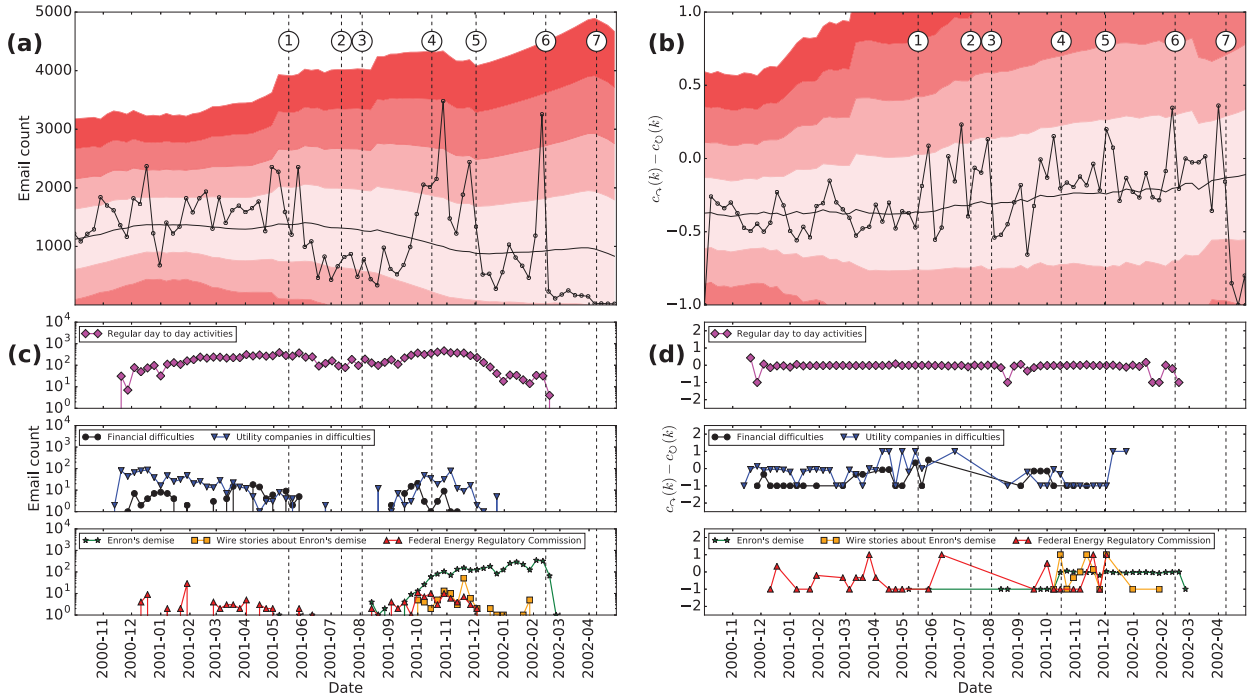

Figure 3.1: Time series of Enron events. (a) Time series of the number of emails. (b) Time series of the difference between the inter- and intra-community link ratios. (c) Time series of the number of emails classified by topics. (d) Time series of the difference between the inter- and intra-community link ratio classified by topics.

Kochen, 1989] to take into account that the dataset is unbalanced [Saito and Rehmsmeier, 2015]. The proposed approach performs generally better than volume-based detection, with noticeable improvements at lower resolutions. For $\mathrm{m} = 7$, the proposed method has a perfect performance.

3.4.2 Boston Marathon

3.4.2.1 Mention Network

As in the previous section, we evaluate whether the communication volume relates to events associated with the Boston Marathon bombing (depicted by the dashed vertical lines and numbers). These events are described in Table 3.2 and have been referenced in [Sutton et al., 2014, Starbird et al., 2014]. Figure 3.4(a) shows the hourly number of English mentions during April 2013. We use the same visualization conventions as in Figure 3.1(a)-(b). Here the window length of the moving average is equivalent to four days of data (96 samples). Note that on 2013-04-08 at 07:00

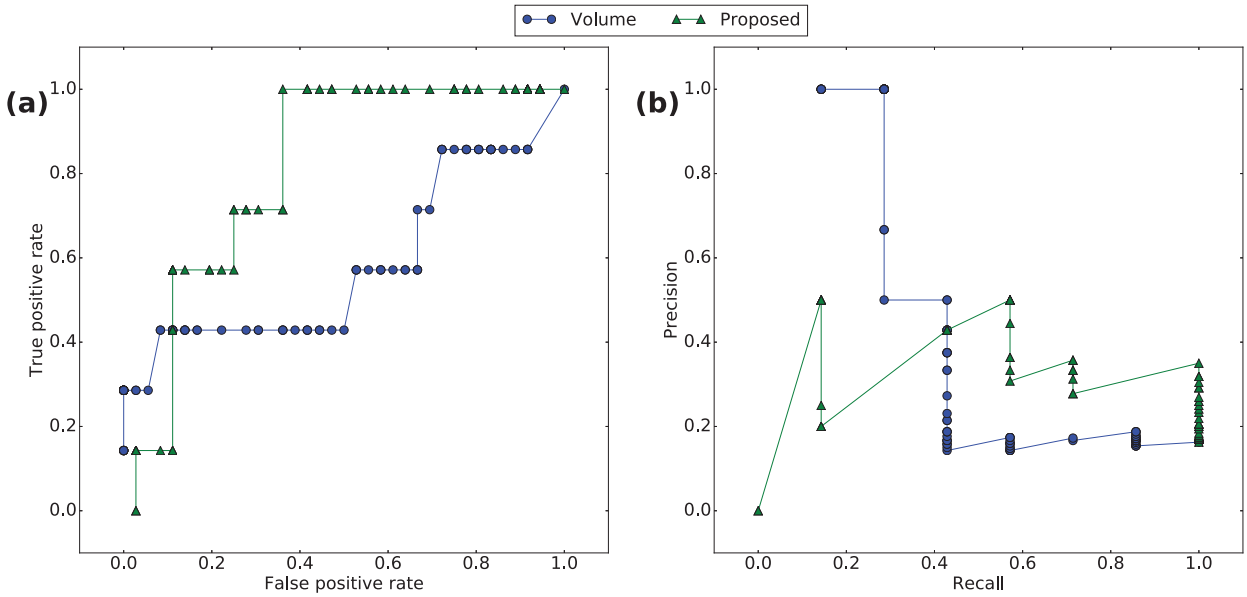

Figure 3.2: Performance comparison for the Enron case when $\mathfrak{m} = 2$ weeks. (a) ROC. (b) PRC.

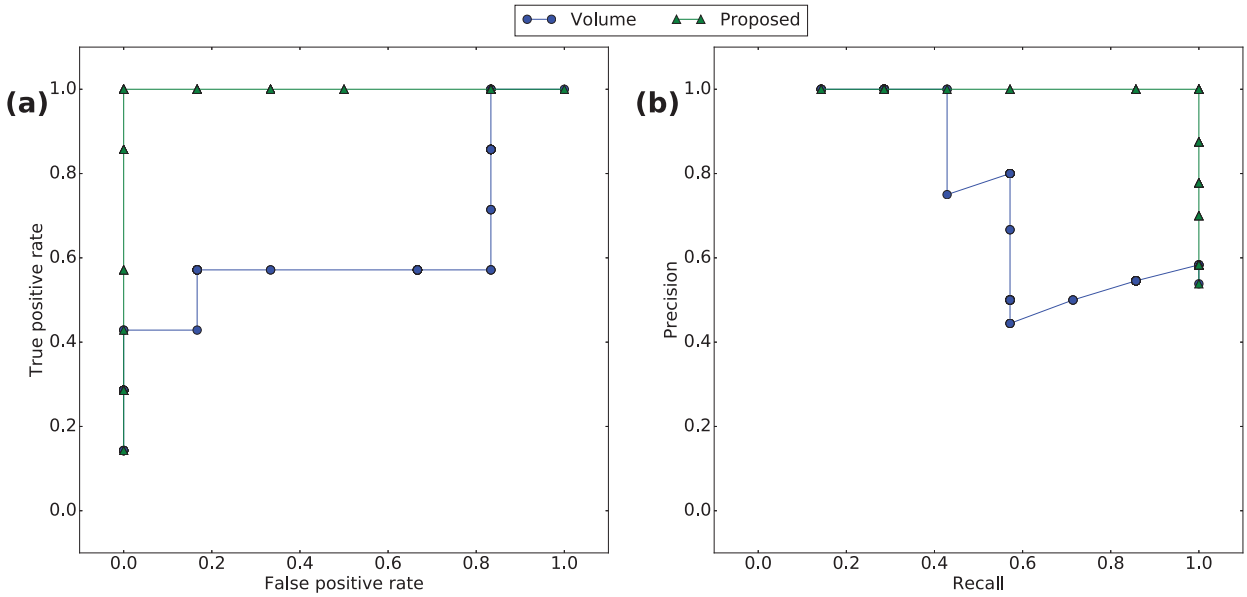

Figure 3.3: Performance comparison for the Enron case when $\mathfrak{m} = 7$ weeks. (a) ROC. (b) PRC.

Event ID	Date	Time	Description
1	2013-04-15	14:49 (UTC)	Bombing
2	2013-04-19	20:42 (UTC)	Firefight and manhunt

Table 3.2: Boston Marathon bombing event description.

UTC, there are some observations that fall three standard deviations beyond the mean. However, these observations are not associated with the events of interest. Similarly, on 2013-04-18, there is a significant decrease in the number of mentions due to missing data. Note also that around events 1 and 2 the number of mentions is comparable to the one in other hours during the observation period, i.e., these data points are statistically insignificant.

Figure 3.4(b) shows the difference between inter- and intra-community link ratios which demonstrates significant deviation at the time of the events (for $\mathfrak{m}_0 = 7$ days). For event 1, the difference moves beyond four standard deviations—suggesting a significant increased in inter- compared to intra-community communications. For event 2, the difference between the ratios is three standard deviations, which is still noticeable compared to other times during the observation period. Figure 3.4(b) also shows other significant deviations. In particular, on 2013-04-21 at 17:00 UTC and 18:00 UTC, the proposed measure falls three standard deviations from the average. This behavior coincides with the hacking of the Associated Press Twitter account on 2013-04-21 around 17:00 UTC. A fake message reported that there had been "two explosions in the white house and Barack Obama [was] injured," which caused financial markets to panic for a few minutes [Jackson, 2013].

We also analyze the contents of the mentions. Figures 3.4(c)-(e) show the distribution of the activity of each hashtag mentioned on 2013-04-15 at different time intervals after the bombing (in EST). In particular, we measure the total number of communications and the difference of the two modalities of communication (inter- and intra communication links). The red cells around the origin indicate that most hashtags are not frequently used. Note also that hashtags tend to be confined inside communities. This is evidenced by the absence of observations with large difference between inter- and intra-communication. Right after the bombing (at 15:00 EST) there are no hashtags with significant difference in the inter- and intra-community level in Fig. 3.4(c). However, in the two subsequent two hour intervals, hashtags related to the bombing emerged distinguished by lots of inter-community communications (see Figs. 3.4(d)-(e)). The highlighted cells correspond to the bombing related hashtags #prayforboston, #BostonMarathon, #PrayForBoston, #Boston and #BostonMararthon. These results demonstrate that the increase in the difference between inter-

and intra-community communication is indeed triggered and driven by the bombing event.

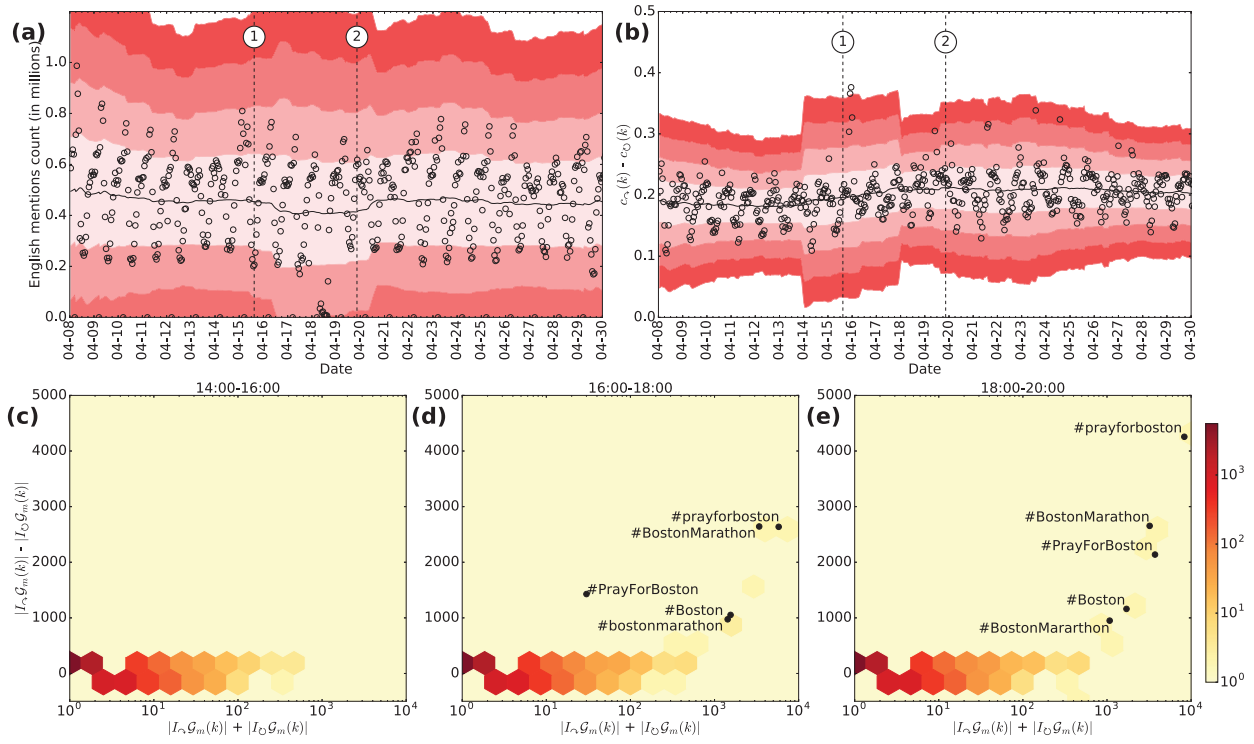

Figure 3.4: Time series analysis of the mention network. (a) Time series of the number of mentions. (b) Time series of the difference between the inter- and intra-community link ratios. (c) Distribution of the number of hashtags based on the total number of links (horizontal axis) and the difference of inter- and intra-community links (vertical axis) during the interval 14:00-16:00 EST on 2013-04-15. (d) Same as (c) during the interval 16:00-18:00 EST. (e) Same as (c) during the interval 18:00-20:00 EST. Hashtags related to the Boston Marathon bombing are highlighted.

3.4.2.2 Retweet Network

Figure 3.5(a) shows the hourly number of English retweets. Note that on 2013-04-08 at 07:00 UTC, there are observations that lay up to three standard deviations from the mean, but are not associated with the events of interest. The spike in the retweet activity is *before* the bombing, which does not relate to the bombing event.

Figure 3.5(b) shows the difference between inter- and intra-community link ratios ($m_0 = 7$ days). For event 1, the difference between community ratios spiked up to four standard deviations, in accordance with Fig. 3.4(b). For event 2, the difference between community ratios is not as

significant as for the case of mentions in Fig. 3.4(b). Other relevant deviations are not observed during the period.

We also explore the content diffused in the retweet network during the hours of the bombing. We measure the distribution of the number of links for each hashtag retweeted on 2013-04-15 at the same time intervals used for Fig. 3.4(c)-(e) and reported them in Fig. 3.5(c)-(e). For the retweet network, we do not observe a significant distribution of hashtags in the vertical axis after the bombing event (see Fig. 3.5(c)). However, Figs. 3.4(d)-(e), Figs. 3.5(d)-(e) show that the bombing related hashtags are placed in regions of low density but with significant difference in inter- and intra-communication links. These hashtags correspond to #prayforboston, #PrayForBoston and #BostonMarathon.

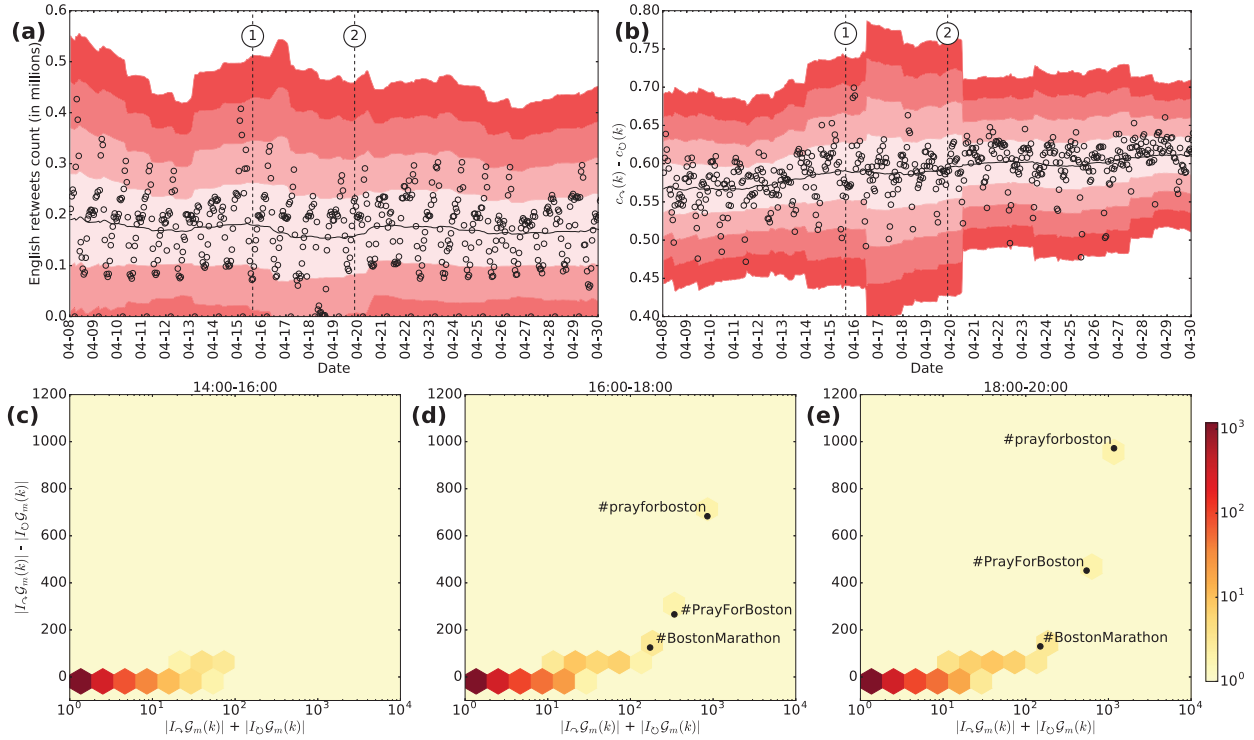

Figure 3.5: Time series analysis of the retweet network. We perform the same analysis as for the mention network and report the number of retweets in (a), the difference between inter- and intra-community link ratios in (b), the distribution of the number of hashtags based on the total number of links (horizontal axis); and the difference of inter- and intra-community links (vertical axis) during the interval 14:00-16:00 EST on 2013-04-15 in (c). (d) Same as (c) during the interval 16:00-18:00 EST. (e) Same as (c) during the interval 18:00-20:00 EST. Hashtags related to the Boston Marathon bombing are highlighted.

3.5 Conclusion

This chapter proposes a community-based method to detect the occurrence of global events in temporal networks. We use the proposed method to detect global events related to the Enron bankruptcy case and the Boston Marathon bombing. In particular, we analyze how the proportion of inter-community links compared to the proportion of intra-community links is a useful signature for identifying the previous described events in each case. Furthermore, we compare the proposed approach with random and volume dependent statistics showing that our method favors precision and recall, that is, we identify the time intervals containing an event, without overlooking their occurrence. Our results suggest that the proposed method filters rich information about the dynamic behavior of temporal networks, which is impossible to observe by measuring the volume of communications. More importantly, patterns of communication across communities are better indicators of the occurrence of events.

Our work has the following limitations. First, the proposed method depends on the definition of an initial community partition. The initial community partition is defined by aggregating the network activity during a fixed period of time (i.e., controlled through m_0). By relying on this strategy, we guarantee that the majority of the users are going to be identified with a community partition and that subsequent interactions in the communication networks can be classified with respect to inter- or intra-community links. However, there is a lot of freedom on the strategy to prepare and update this "normal" community structure. Second, to make the decision on whether a particular data sample includes an event, our reasoning is based on the distance of the observation with respect the moving average of the measure. We implement this criterion by accounting for the number of moving standard deviations (i.e., controlled through δ in Algorithm 1). Clearly, defining how many standard deviations are needed to establish a detection threshold depends on many factors. In the two case studies, the frequency of the formation of the networks is one week for Enron and one hour for the Twitter datasets. Third, in evaluating the performance of the algorithms, the detection intervals are assumed to be proportional to the network formation

intervals. This means that even when an interval is reported to contain an event, there is no notion of temporality with respect to the closeness of the occurrence of the event within that interval. This might be balanced by increasing the length of detection intervals. However, given the length of the observation periods, a limited detection resolution will decrease the performance of the proposed method.

4 Insider Threat Modeling[1]

"If you want to do evil, science provides the most powerful weapons to do evil; but equally, if you want to do good, science puts into your hands the most powerful tools to do so."

— Richard Dawkins

4.1 Introduction

In this chapter, we propose a method for analyzing insider threat actions as anomalies. To that end, we model user-system interactions as a bipartite graph. Our method is based is based on capturing the regular committing behavior of developer's inside communities. In particular, we quantitatively show that the committing behavior of developers changes after an external (precipitating) event is announced. We capture this signature by quantifying the proportion of commits inside as opposed to outside a community. Our study is based on analyzing the committing behavior of developers into code repositories using a Cisco codebase and verified internal events.

4.2 Problem

Insiders are employees that must be trusted with access to sensitive information, and because of that trust can be a major threat. Insiders have compromised organizations in multiple domains including manufacturing [Reuters, 2011], finance [FBI, 2010], government [Edwards and Hoosenball, 2016], and even scientific research [Culp, 2013]. Even worse, insiders attacks are consistently

[1]The content of this chapter was initially published as a conference paper [Moriano et al., 2017b] and then an extended version of it was published as a journal article [Moriano et al., 2018a] in collaboration with Jared Pendleton, Steven Rich, and L. Jean Camp. Pablo Moriano is the primary researcher on both works and made all the analysis and figures therein.

catalogued as the most costly given the elevated privilege that insiders have in terms of trust and access [Ponemon Institute, 2016]. This makes the insider issue one of the most challenging problems in computer security [Bishop et al., 2014].

As with many other complex systems (e.g., the Internet, online social networks, and the brain), information systems consist of a large number of interacting elements (e.g., users, services, devices, files) in which the aggregate activity of the system cannot be derived by analyzing individual contributions, i.e., their aggregate behavior is nonlinear. Graphs, where nodes represent the elements and edges capture the interactions between the elements of the system, have been used across multiple domains to capture the interactions between the elements of complex systems [Vespignani, 2009, Newman, 2010]. The use of graphs to study the structure of complex systems has revealed some plausible explanations for the emergence of collective behavior in these systems such as the understanding of regular and anomalous behavior [Akoglu et al., 2015]. In this work, we treat the malicious insider as an anomaly and use bipartite graphs to detect their anomalous behaviors.

The resulting focus on malicious patterns, as opposed to malicious nodes, implements an assumption that the malicious insider is not intrinsically hostile. Rather, malicious behaviors can emerge over time or in respect to specific conditions. Static graphical analysis is based on the analysis of graph snapshots and cannot integrate temporal patterns. In contrast, the study of temporal graphs, where information of single graph snapshots is aggregated, tends to reflect more accurately the evolution of the system as nodes and edges appear and disappear over time [Ranshous et al., 2015, Holme and Saramäki, 2012]. The focus of this work is to understand the malicious behaviors over time rather than identifying the static malicious nodes.

To understand such complex systems, empirical data with detailed temporal information is a prerequisite. Correct temporal information is much more readily available as a source of ground truth than correctly labeled insider threat datasets. In the context of information systems, temporally annotated datasets are widely available thanks to the presence of user-system interaction logs. This enables the use of graph mining analytics for the understanding of anomalous behavior such

as the one that insiders might pose [Eberle et al., 2010, Parveen et al., 2011].

In this chapter, we characterize and detect anomalous events in an information system based on a centralized version control system[2]. We identify time intervals during which significant changes in the structure of the temporal graphs may correspond to functional change points, e.g., a precipitating event[3]. This problem has also been referred to as change point detection [Barnett and Onnela, 2016].

We model user-system interactions in a version control system as a temporal bipartite graph where interactions occur exclusively between two types of nodes, (i) users and (ii) software components[4]. Note that the edges in this graph are only between these two types of nodes [Heymann and Le Grand, 2013]. A one-mode projection of this graph is the *user graph* in which two nodes (users) are connected if they have interacted at least once with the same component [Zhou et al., 2007]. Our methodology includes studying the evolution of the one-mode user graph to identify topological properties that characterize the system's normal behavior. Among these observed properties, those that do not follow the norm of the regular pattern are assumed to indicate the presence of an anomalous event. Such an event may indicate a potential insider incident or, at least, an event that requires further investigation [Rashid et al., 2016].

In particular, the user graph allows us to explore the impact of precipitating events in user-system interactions [Nurse et al., 2014]. Precipitating events are key events that have the potential to trigger insiders to become a threat to their employer. We hypothesized that precipitating events impact the behavior of interactions between users and components in the version control system

[2]A centralized version control system keeps the history of changes on a central server from which everyone requests the latest version of the work and pushes the latest changes to, e.g., Concurrent Versions System and IBM Rational ClearCase.

[3]A precipitating event corresponds to a large-scale event that causes concerning behaviors in employees and may result in otherwise trustworthy employees becoming threats or taking risks that increase the insider threat. In this category, we include layoffs, significant restructuring, and plant or facility closure. The term was first used in the insider threat literature in Moore et al. [Moore et al., 2008].

[4]A software component is a software module that encapsulates a set of related functions or data, and it is part of a larger software system. For example, the TCP/IP software component of an operative system. Hereafter, we refer to software components as simply components.

by changing patterns of committing behavior. To test this hypothesis, we model and compare the volume of interactions between users over similar or related components as opposed to non-related components over time. To capture sets of users with similar patterns of interaction, we rely on the notion of community structure to identify communities, or clusters, i.e., groups of nodes having higher probability of being connected to each other than to members of other groups [Fortunato and Hric, 2016]. We show that the volume of interactions between users that contribute to unrelated components increases when precipitating events are announced. This indicates the impact of precipitating events in increasing the likelihood of a change in the interacting behavior between users and components, which might be a signal to monitor before an insider attack is committed.

4.3 Methods

Our method builds graphs of user-system interactions and uses these to identify anomalous patterns. Anomalies are identified when engineers interact with multiple components where there is no history of interaction, particularly where none of their team members are interacting or have a history of interaction with those components. Performance is measured by the ability of the algorithm to detect increases in anomalous behavior after precipitating events.

4.3.1 Temporal Abstraction

Consider the sequence of n intervals $A = \{A_1, A_2, \ldots, A_n\} = \{A_k\}_{k=1}^{n}$, where

1. $A_k = [a_k, a'_k)$ for all $k < n$ and $A_n = [a_n, a'_n]$ for $k = n$;
2. $a_k < a'_k = a_{k+1}$ for all k; and
3. $a'_k - a_k = a'_\ell - a_\ell$ for all k, ℓ

An interval represents a fixed-length unit of time, e.g., a day of data. Condition (1) implies that all intervals are left-closed and right-open (except the last one which includes a'_n). It guarantees that the sequence of intervals is disjoint. Condition (2) implies that intervals are non-empty. Note that

a'_k and a_{k+1} represent the time instants of a transition between intervals. For any interval A_k, the right endpoint a'_k corresponds to the left endpoint of the interval A_{k+1}. Together with Condition (1), Condition (2) guarantees that the union of all intervals $\bigcup_{k=1}^{n} A_k = [a_1, a'_n]$ is a closed interval. Finally, Condition (3) requires that any two intervals are of equal length.

4.3.2 Bipartite Graph Abstraction

A bipartite graph is a graph with two types of nodes. One type of node represents the original nodes (top nodes), while the other represents the groups with which they interact (bottom nodes) [Guillaume and Latapy, 2004].

Let $\mathcal{H}_\top$ be the set of top nodes (e.g., the set of engineers). Similarly, let $\mathcal{H}_\bot$ be the set of bottom nodes (e.g., the set of components). Note that $\mathcal{H}_\top$ and $\mathcal{H}_\bot$ are disjoint sets of nodes. Furthermore, let $\mathcal{V}(k) \subseteq \mathcal{H}_\top \cup \mathcal{H}_\bot$ be the subset of nodes that interact (i.e., engineers and components) during interval $A_k = [a_k, a'_k)$. Let $\mathcal{W}(k) = \{\Omega_{ij}(k) : (i,j) \subseteq \mathcal{H}_\top \times \mathcal{H}_\bot\}$ be the incidence matrix of weights $\Omega_{ij}(k)$ that captures the number of interactions between node i and node j during interval A_k. Let $\mathcal{G}(k) = (\mathcal{V}(k), \mathcal{W}(k))$ represent a weighted bipartite graph that captures all interactions that occur from endpoints a_k to a'_k, $k \in \{1, 2, \ldots, n\}$. Note that we do not differentiate between dynamics within an interval. The sequence $\{\mathcal{G}(k)\}_{k=1}^{n}$ denotes the bipartite graph series G.

4.3.3 One-Mode Projection Abstraction

Bipartite graphs can be projected to one-mode projection graphs (with nodes of just one type). Let $\mathcal{G}_\top(k) = (\mathcal{H}_\top(k), \mathcal{W}_\top(k))$ be the top projection of $\mathcal{G}(k)$. Two nodes of $\mathcal{H}_\top(k)$ are connected if they have at least one neighbor in common in $\mathcal{G}(k)$, i.e., $\mathcal{W}_\top(k) = \{\omega_{uv}(k) : u, v \subseteq \mathcal{H}_\top\}$, where

$$\omega_{uv}(k) = \sum_{r=1}^{|\mathcal{H}_\bot|} \Omega_{ur}(k) + \Omega_{vr}(k)$$

The sequence $\{\mathcal{G}_\top(k)\}_{k=1}^{n}$ denotes the top one-mode projection graph series $G_\top$. Correspondingly, the bottom projection $\mathcal{G}_\bot(k) = (\mathcal{H}_\bot(k), \mathcal{W}_\bot(k))$ is defined dually as it is illustrated in Figure 4.1. The sequence $\{\mathcal{G}_\bot(k)\}_{k=1}^{n}$ denotes the bottom one-mode projection graph series $G_\bot$. In the rest

of this chapter, we devote our study in terms of $G_\top$ which is the one-mode projection graph of user-system interactions, i.e., the projection in which nodes are exclusively engineers.

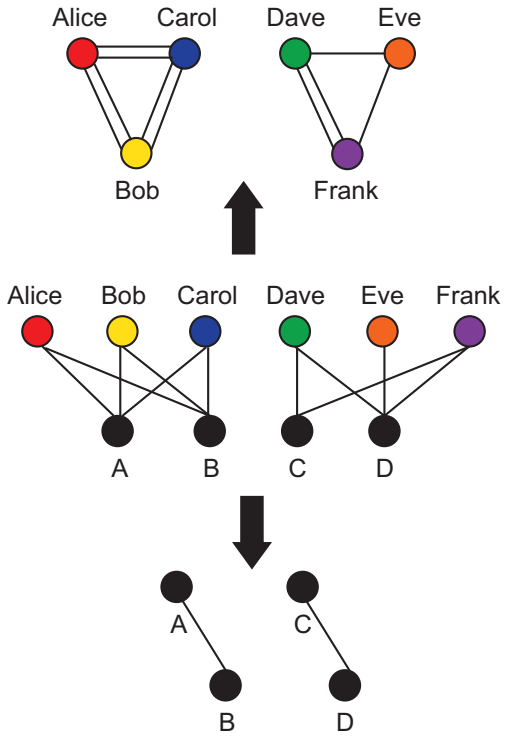

Figure 4.1: Bipartite graph abstraction. The top panel represents the engineer projection. The middle panel represents the original bipartite graph. The bottom panel represents the component projection.

4.3.4 Detection Problem

We use $G_\top$, which captures the dynamics across intervals A_k, $k \in \{1, 2, \ldots, n\}$, as the basis for defining the anomaly event detection problem. In doing so, we evaluate the outcomes of anomaly detection by measuring structural properties with respect to the cumulative one-mode graph segment of length $m \in \mathbb{Z}^+$ defined as

$$\begin{aligned} \mathcal{G}_\top^m(k) &= (\mathcal{V}_\top^m(k), \mathcal{W}_\top^m(k)) \\ &= \bigoplus_{k'=k-m+1}^{k} \mathcal{G}_\top(k') = \mathcal{G}_\top(k-m+1) \oplus \cdots \oplus \mathcal{G}_\top(k) \end{aligned}$$

where

$$\mathcal{V}_{\top}^{m}(k) = \bigcup_{k'=k-m+1}^{k} \mathcal{V}_{\top}(k') \text{ and } \mathcal{W}_{\top}^{m}(k) = \sum_{k'=k-m+1}^{k} \mathcal{W}_{\top}(k')$$

For example, if $m = 7$, we aggregate data to form weekly graph segments.

Let lm, where $l \in \mathbb{Z}^{+}$ represents the smallest interval at which we evaluate the outcomes of anomalous detection (called the detection resolution). Note that if $l > 1$, then the intervals at which the graph segments are evaluated are not the same as the ones at which they are formed. The finest detection granularity satisfies $l = 1$, i.e., when the detection resolution is the same as the graph segment formation intervals. A larger value of l reflects that anomalous events are captured by the aggregation of consecutive graph segments. For instance, if $l = 2$, then an algorithm for detection aims to determine whether such an event occurs within intervals $(a_{k-lm+1}, a'_k] = (a_{k-2m+1}, a'_k]$, $k \in \{2m, \ldots, n\}$. Let $\bar{n} = \left\lfloor \frac{n}{lm} \right\rfloor$ be the total number of times the algorithm with resolution lm has to decide whether an event occurs. Let the set $E \subseteq \{1, 2, \ldots, \bar{n}\}$ represent the intervals at which at least one event occurs. The detection problem is specified as follows.

Given:

(i) A one-mode projection graph series $G_{\top} = \{\mathcal{G}_{\top}(k)\}_{k=1}^{n}$; and

(ii) A detection resolution $1 \leqslant lm < n$.

We want to:

(ii) Design a detection algorithm that identifies the subset of intervals $\hat{E} \subseteq E$ in which at least one anomalous event occurs.

Condition (i) requires that the dataset can be modeled as a series of one-mode projection graphs that aggregate the interactions occurring during each interval. Condition (ii) assumes that a parameter can be selected to enable detection of anomalies at a desired timescale.

4.3.5 Algorithm Performance Abstraction

Consider a sequence of detection intervals $B = \{B_1, B_2, \ldots, B_{\bar{n}}\} = \{(a_{(t-1)lm+1}, a'_{tlm}]\}_{t=1}^{\bar{n}} = \{B_t\}_{t=1}^{\bar{n}}$. To measure performance, the output of the detection algorithm $\hat{E}$ is mapped into the sequence of intervals B. Let $\hat{e} \in \hat{E}$ be the index of a detection interval that is denoted as anomalous by the detection algorithm (i.e., the algorithm indicates the occurrence of at least one anomalous event within the interval). The set $\hat{E}$ can be represented by the indicator vector

$$\hat{O} = \bigvee \{\mathbb{1}_{B_t}(lm\hat{e}),\ \forall t \in \{1, 2, \ldots, \bar{n}\}\},\ \forall \hat{e} \in \hat{E}$$

where $\bigvee$ represents the OR operator and $\mathbb{1}_{B_t}(lm\hat{e})$ denotes the indicator function

$$\mathbb{1}_{B_t}(lm\hat{e}) = \begin{cases} 1 & \text{if } lm\hat{e} \in B_t \\ 0 & \text{if } lm\hat{e} \notin B_t \end{cases}$$

In other words, if $\mathbb{1}_{B_t}(lm\hat{e}) = 1$, the algorithm identifies an anomalous event in the detection interval $(a_{(t-1)lm+1}, a'_{tlm}]$ and labels it as an anomalous interval. The indicator vector $\hat{O}$ describes the interval indices, i.e., $t \in \{1, 2, \ldots, \bar{n}\}$ that contain an anomalous event.

Moreover, to characterize the occurrence of actual events during an interval, we define $e \in E$ as the index of a detection interval that is anomalous based on the ground truth. Let the indicator vector $O = \bigvee \{\mathbb{1}_{B_t}(lme),\ \forall t \in \{1, 2, \ldots, \bar{n}\}\},\ \forall e \in E$ represents the intervals that are anomalous based on the ground truth, i.e., the distribution of the anomalous events over the set of the $\bar{n}$ detection intervals. Figure 4.2 illustrates the proposed modeling framework. For example, suppose that $E = \{s, \bar{n}\}$ (represented by the horizontal arrows) and $\hat{E} = \{s\}$ (represented by the horizontal crossed arrow). To pinpoint the detection interval s, there might exist a time index $r = ms$ such that $\mathbb{1}_{B_s}(r) = 1$. This is represented by the vertical arrows in Figure 4.2. This is also illustrated in Figure 4.3. Here it is possible to see the correspondence between the intervals of formation of the cumulative graph segment and the segments at which the algorithm is evaluated.

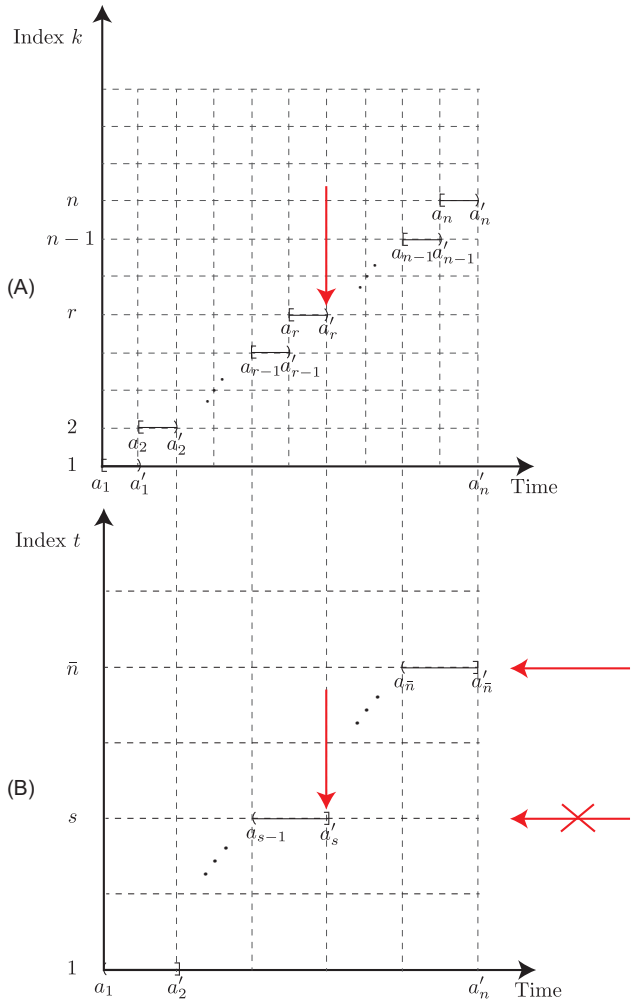

Figure 4.2: Abstraction of the detection problem. The top panel refers to the sequence of intervals that are used to build the graphs (here the graph formation interval $\mathtt{m} = 1$). The bottom panel illustrates the aggregation of intervals to evaluate the performance of the detection algorithm (here detection resolution $\mathtt{lm} = 2$). The vertical arrows represent the location of an anomalous event in both temporal representations. The horizontal arrows illustrate the sets $\mathtt{E}$ and $\hat{\mathtt{E}}$.

4.3.6 Algorithm Performance Measure

The performance of a detection algorithm is measured based on identifying the anomalous detection intervals. Specifically, the performance of an algorithm is specified based on the set of time intervals $\hat{\mathtt{E}}$ reported as anomalous by the detection algorithm and the set of time intervals $\mathtt{E}$ in which anomalies occur (ground truth).

We compare the performance of the detection algorithms using the true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) of the detection results. In particular,

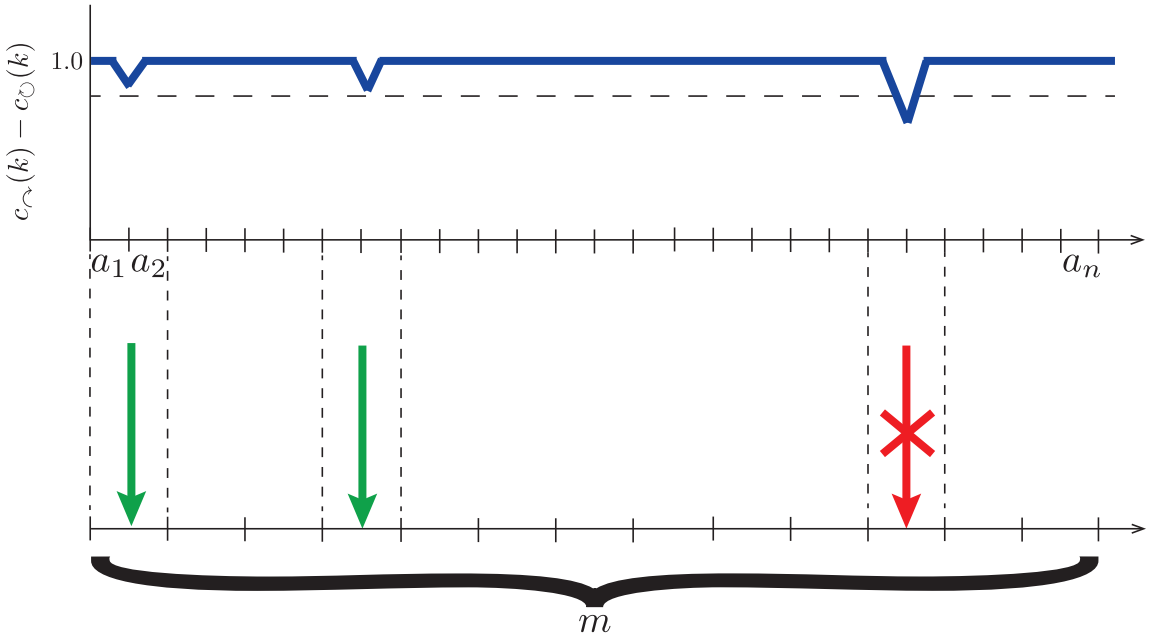

Figure 4.3: Correspondence between the graph formation intervals and the intervals at which the algorithm is evaluated. Here m is the number of intervals at which the algorithm is evaluated.

$\mathrm{TP} = \mathrm{O} \cdot \hat{\mathrm{O}}$, $\mathrm{FP} = \mathrm{O}' \cdot \hat{\mathrm{O}}$, $\mathrm{FN} = \mathrm{O} \cdot \hat{\mathrm{O}}'$, and $\mathrm{TN} = \mathrm{O}' \cdot \hat{\mathrm{O}}'$ where the symbol "$\cdot$" represents the dot product between two vectors, and O' and $\hat{\mathrm{O}}'$ represents the complement of O and $\hat{\mathrm{O}}$ respectively.

In other words, a detection algorithm specifies the intervals based on a detection criterion. Similarly, to measure performance, it is necessary to know the ground truth anomalous events. The detailed pseudo-code for the algorithm's performance measure is presented in Algorithm 3. Next, we introduce a detection criterion based on the dynamics of the formation of communities and the interaction of engineers across and within them.

4.3.7 Proposed Algorithm

The proposed algorithm defines detection signatures based on deviations from the regular process of community interaction. To do so, we explore whether variations in the number of edges across communities (with respect to the total number) are indicators of anomalous events. This is done by comparing edges in the user graph with respect to a community partition reference over aggregate

Algorithm 3 Algorithm-Performance ($\hat{E}$, E, $\bar{n}$)

$\hat{O} \leftarrow zeros(\bar{n})$
for $\hat{e} \in \hat{E}$ **do**
  $\hat{O}_{\hat{e}} \leftarrow \{\}$
  for $t \in \{1, 2, \ldots, \bar{n}\}$ **do**
    $\hat{O}_{\hat{e}} \leftarrow \hat{O}_{\hat{e}} \cup \mathbb{1}_{B_t}(\hat{e})$
  end for
  $\hat{O} \leftarrow \hat{O}$ OR $\hat{O}_{\hat{e}}$ (element wise)
end for
$O \leftarrow zeros(\bar{n})$
for $e \in E$ **do**
  $O_e \leftarrow \{\}$
  for $t \in \{1, 2, \ldots, \bar{n}\}$ **do**
    $O_e \leftarrow O_e \cup \mathbb{1}_{B_t}(e)$
  end for
  $O \leftarrow O$ OR O_e (element wise)
end for
$O' \leftarrow$ NOT (O)
$\hat{O}' \leftarrow$ NOT $(\hat{O})$
$TP \leftarrow O \cdot \hat{O}$
$FP \leftarrow O' \cdot \hat{O}$
$FN \leftarrow O \cdot \hat{O}'$
$TN \leftarrow O' \cdot \hat{O}'$
return (TP, FP, FN, TN)

data.

Let the initial cumulative one-mode graph segment of length m_0, $1 \ll m_0 \ll n$ be defined as

$$\begin{aligned} \mathcal{G}_\top^{m_0} &= (\mathcal{V}_\top^{m_0}, \mathcal{W}_\top^{m_0}) \\ &= \bigoplus_{k'=1}^{m_0} \mathcal{G}_\top(k') = \mathcal{G}_\top(1) \oplus \cdots \oplus \mathcal{G}_\top(m_0) \end{aligned}$$

where $\mathcal{V}_\top^{m_0} = \bigcup_{k'=1}^{m_0} \mathcal{V}_\top(k')$ and $\mathcal{W}_\top^{m_0} = \sum_{k'=1}^{m_0} \mathcal{W}_\top(k')$.

The proposed detection algorithm requires the following assumption:

(A1) The initial cumulative graph segment $\mathcal{G}_\top^{m_0}$ can be naturally divided in non-overlapping communities, i.e., groups of nodes that can be grouped into subsets such that each set of nodes is densely connected internally and in which nodes belong to a single group [Newman, 2004].

Let the set $T = \{m_0 + m, m_0 + 2m, \ldots, \bar{n}m\}$ captures the time intervals at which the algorithm will be applied. Note that for $k \in T$, the series $\{\mathcal{G}_\top^m(k)\}$ forms a set of non-overlapping cumulative graph segments. The proposed algorithm pinpoints anomalous events by measuring the proportions of inter- and intra-community edges of the graph $\mathcal{G}_\top^m(k)$ with respect to the community partition of

$\mathcal{G}_{\top}^{m_0}$, i.e., we want to identify the set $\hat{E}$ based on the diversification of community edges. Figure 4.4 shows a characterization of that situation.

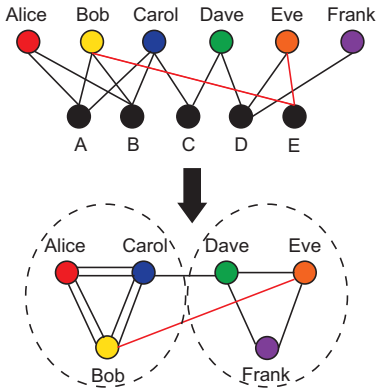

Figure 4.4: Malicious activity as identified in the bipartite graph. The top panel represents an unusual interaction between Bob and Eve with component "E." The bottom panel represents the corresponding one mode projection graph with the anomalous edge crossing communities.

To do so, let $C(\mathcal{G}_{\top}^{m_0}) = \{0, 1, \ldots, c\}$ be a set of unique community identifiers where $c + 1$ is the total number of detected communities in the initial cumulative graph segment $\mathcal{G}_{\top}^{m_0}$. The community to which engineer $i \in \mathcal{V}_{\top}^{m}(k) \cap \mathcal{V}_{\top}^{m_0}$ is assigned (with respect to $\mathcal{G}_{\top}^{m_0}$) is given by $c_i(k) : i \to C(\mathcal{G}_{\top}^{m_0})$. We computed the community partition of the initial cumulative graph segment using the Infomap algorithm [Rosvall and Bergstrom, 2008]. Following similar ideas as in [Weng et al., 2013], let the set of inter-community edges be $I_{\frown}(\mathcal{G}_{\top}^{m}(k)) = \{(u, v) : \omega_{uv}(k) > 0 \wedge (c_u(k) \cap c_v(k)) = \varnothing\}$ and intra-community edges be $I_{\circlearrowleft}(\mathcal{G}_{\top}^{m}(k)) = \{(u, v) : \omega_{uv}(k) > 0 \wedge (c_u(k) \cap c_v(k)) \neq \varnothing\}$. We also define the inter- and intra-community ratio as

$$c_{\frown}^{m}(k) = \frac{|I_{\frown}(\mathcal{G}_{\top}^{m}(k)|}{|I_{\frown}(\mathcal{G}_{\top}^{m}(k))| + |I_{\circlearrowleft}(\mathcal{G}_{\top}^{m}(k))|} \tag{4.1}$$

$$c_{\circlearrowleft}^{m}(k) = \frac{|I_{\circlearrowleft}(\mathcal{G}_{\top}^{m}(k)|}{|I_{\frown}(\mathcal{G}_{\top}^{m}(k))| + |I_{\circlearrowleft}(\mathcal{G}_{\top}^{m}(k))|} \tag{4.2}$$

respectively.

In particular, we are interested in identifying time intervals k, where $c_{\circlearrowleft}^{m}(k) - c_{\frown}^{m}(k)$ is below median-3σ or above median+3σ. The median is used instead of the mean because this measure (over the entire period of study) cannot be assumed follow a normal distribution since appropriate hypothesis testing demonstrates that the normal distribution is not a good candidate to model the

generation of the empirical observations. We used the interquartile range to estimate σ as it has been studied by others, e.g., [Koutra et al., 2013]. The detailed pseudo-code for this algorithm is shown in Algorithm 4.

Algorithm 4 Event-Detection ($G_\top$, m_0, m)

Compute community partition of $\mathcal{G}_\top^{m_0}$
$Y \leftarrow \{\}$ ▷ Array of intra−inter ratio samples
for k in $\{m_0 + m, m_0 + 2m, \ldots, \bar{n}m\}$ **do**
    Build $\mathcal{G}_\top^m(k) = \bigoplus_{k'=k-m+1}^{k} \mathcal{G}_\top(k')$
    Compute $I_\curvearrowright(\mathcal{G}_\top^m(k))$
    Compute $I_\circlearrowright(\mathcal{G}_\top^m(k))$
    Calculate $c_\curvearrowright^m(k)$ and $c_\circlearrowright^m(k)$ using eqs. 4.1 and 4.2
    $Y \leftarrow Y \cup \{c_\circlearrowright^m(k) - c_\curvearrowright^m(k)\}$
end for
median $\leftarrow \hat{F}_Y^{-1}(0.50)$ ▷ $\hat{F}$ means the empirical CDF
$\delta \leftarrow \hat{F}_Y^{-1}(0.75) - \hat{F}_Y^{-1}(0.25)$ ▷ The interquartile range
$\hat{E} \leftarrow \{\}$
for k in $\{m_0 + m, m_0 + 2m, \ldots, \bar{n}m\}$ **do**
    if $Y(k) <= (\text{median} - 3\sigma)$ or $Y(k) >= (\text{median} + 3\sigma)$ **then**
        $\hat{E} \leftarrow \hat{E} \cup \{k\}$
    end if
end for
return $\hat{E}$

In order to compare the performance of the algorithms, we replace the computation of $c_\circlearrowright^m(k) - c_\curvearrowright^m(k)$ by the respective graph topological property, e.g., nodes, edges, connected components, average degree, maximum degree, or maximum weight with respect to $\mathcal{G}_\top^m(k)$.

4.3.8 Dataset

IBM Rational ClearCase (hereafter ClearCase) is an enterprise-grade software configuration management system. Among its main features, it provides version control functionalities to large- and medium-size organizations allowing them to track software projects with thousands of developers. As of the date of this writing, ClearCase has a market share of about 2.5% among software configuration management competitors with 55% of their customers in the U.S. [iDatalabs, 2017].

The ClearCase dataset analyzed in this chapter comprises the complete activity between engineers and components in a major computer software enterprise. Components are software packages that encapsulate a set of related functions and store metadata allowing version control. In particular, we used data that spans 22 years from May 4, 1992 to March 23, 2014. We extracted the data

from the source code base management database. Instances with no reference to the engineer or component name were not taken into account in this analysis. These comprised a negligible percentage of instances, i.e., on the order of 8×10^{-6}. (Thus the number of interactions in this dataset that were not captured in our graph mining method is sufficiently small that manual examination for insider activity would be quite feasible.)

Using this dataset, we built bipartite graphs to capture the interactions between engineers and components. In this bipartite graph, nodes are represented exclusively by engineers and components. Edges in the bipartite graph represent interactions, i.e., any type of activity that engineers have with components, including: commit a file, create a file, delete a file, create a branch, tag a branch, sync a branch, and collapse a branch. We did not differentiate between these different interactions and treat them as the same type of edges.

The dataset comprises $10,253$ distinct engineers, $1,729$ distinct components, and $12,577,667$ interactions during the observation period. Remember that our hypothesis is grounded on the idea that precipitating events might lead to structural changes in the committing behavior of engineers. With that in mind, Table 4.1 summarizes the details of the incidents used in this study, i.e., precipitating events that were announced and validated internally by the enterprise. These events correspond to limited restructuring events and had and effect in all business units of the enterprise.

Table 4.1: Summary of precipitating events during the observation period.

Event ID	**Date**	**Jobs impacted**	% **affected employees**
(1)	2001-04-16	8500	22.4
(2)	2011-07-18	6500	9.1
(3)	2012-07-23	1300	1.9
(4)	2013-03-26	500	0.7
(5)	2013-08-09	4000	5.3

4.4 Results

In this section, we present the results of the analyses on the bipartite graphs and the one-mode projection (or user graph) of user-system interactions. In the following analysis, our unit of time reference is the day, i.e., the scale of the variable $\mathtt{k}$. To estimate the length of the window $\mathtt{m}$ (the window length that we use to accumulate interactions among engineers), we relied on the methodology proposed by [Benamara and Magnien, 2010], which estimated that the size of an observable window for a rigorous characterization of graph properties is at least one week, i.e., $\mathtt{m} = 7$ days. This means that we build the bipartite and one-mode projection graphs by aggregating data over non-overlapping windows of 7 days (every week starting on Monday).

We compare the results of the proposed event detector framework to random chance. The purpose of this comparison is to ensure that the phenomena we identify are not a result of noise or simply the result of having stochastic data. We then compare our approach with metrics that are based on the volume of interactions. That is, we test if the proposed approach identifies insider risk more accurately than those that identify employees by frequency or intensity of access. Sheer counts of access are a core component of risk-based or accounting-based insider threat approaches. The model proposed here is more accurate and more precise. The model also offers fewer false negatives (i.e., higher recall).

We used the same visualization conventions for every plot. The blue solid lines show the raw data. Recall that the raw data corresponds to the empirical measures for each graph topological property. Dashed black lines represent the dates of the precipitating events listed in Table 4.1 with their corresponding label in a circle.

The results of these show that the precipitating events cannot be distinguished from other events using simple graph-based statistics. Our assumption is that although individual events, such as economic stress, may result in an individual becoming an insider threat, only systematic organizational changes should be correlated with large-scale increases in insider threat behaviors.

In the following two sections we report basic graph properties. The purpose of this is to illus-

trate that naive application of graph mining on bipartite graphs without inclusion of community dynamics is an inadequate indicator of insider threats.

4.4.1 Bipartite Graph Properties Series

We report measures related to the number of nodes, edges, connected components, average degree, maximum degree, and maximum weight for the bipartite graphs. The specific properties that we measured from these graphs are listed here for the reader. Note that edges only capture interactions between engineers and components. The number of nodes is $|\mathcal{V}(k)|$, which is the total number of engineers and components. The degree of node i is $d_i(k)$, i.e., its number of neighbors. The degree of a node is either the number of components that are touched by a single engineer or the number of engineers that touch a component. The set of edges of the graph $\mathcal{G}(k)$ is $\mathcal{E}(k)$ is the total number of unique component/engineer interactions. The number of edges is $|\mathcal{E}(k)| = \sum_{i \in \mathcal{V}(k)} d_i(k)/2$, which means this represents total number of interactions in the bipartite graph, i.e., system activity. A connected component is a subgraph in which any two nodes are connected to each other by paths. This means subgraphs can be the result of connections or similarities in connections between engineers and/or components so that any two vertices are connected by any path of interactions. The average degree of graph $\mathcal{G}(k)$ is $2 \times |\mathcal{E}(k)|/|\mathcal{V}(k)|$. This means the average number of interactions that nodes have reflects the fact that there are two entities in an interaction. The maximum degree of graph $\mathcal{G}(k)$ is the maximum number of neighbors in the graph, i.e., $\max\{d_i(k), \forall i \in \mathcal{V}(k)\}$. This is the maximum number of interactions among the nodes (either with engineers or components). The maximum weight of a graph $\mathcal{G}(k)$ is the maximum weight among the edges in the graph, i.e., $\max\{\omega_{ij}(k), \forall i, j \in \mathcal{V}(k)\}$. This means the maximum weight value among the interactions, i.e., the interaction with the higher intensity.

Figure 4.5 (top) shows the observed number of nodes (i.e., engineers and components) in the bipartite graphs. Figure 4.5 (middle) shows the number of unique edges representing the number of interactions between engineers and/or components. Figure 4.5 (bottom) shows the number of connected components in the bipartite graphs. There is a constant increase in the number of

nodes since approximately 2002 because there is an increase of both engineers and components. The tendency starts to decrease after approximately 2010 as other version control systems were adopted. Thus, after 2010, the data are a large sample rather than a comprehensive dataset. This tendency is also reflected in the number of edges of the bipartite graphs, which is correlated with the number of nodes [Leskovec et al., 2005]. The movement of some core technologies to a different versioning system is reinforced by the continuous increase in the number of connected components in the bipartite graph, which indicates a less integrated core of components.

Similarly, Figure 4.6 shows the time series of average degree, maximum degree, and maximum weight for the bipartite graphs. Note that under the presence of spikes in the measured signal, they are not correlated with the vertical lines that indicate the occurrence of the precipitating events.

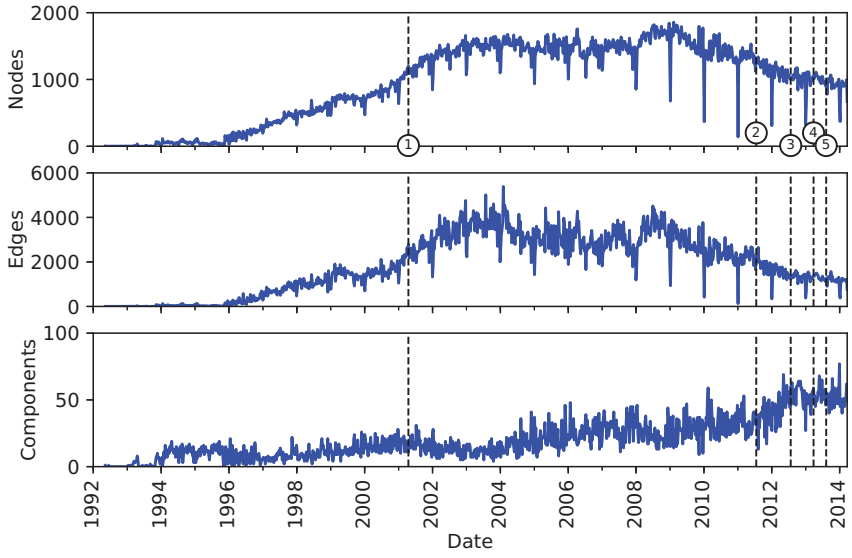

Figure 4.5: Time series of the number of nodes (top panel), edges (middle panel), and connected components (bottom panel) for the bipartite graphs.

4.4.2 One-Mode Projection Graph Properties series

Here, we report results on the same properties as we did for the bipartite graphs, i.e., nodes, edges, components, average degree, maximum degree, and maximum weight. Note that, in this case, edges occur when two engineers have interact with the same component (i.e., same code repository) based on the bipartite one-mode projection. The number of nodes is $|\mathcal{H}_{\top}(\mathtt{k})|$. This means the total

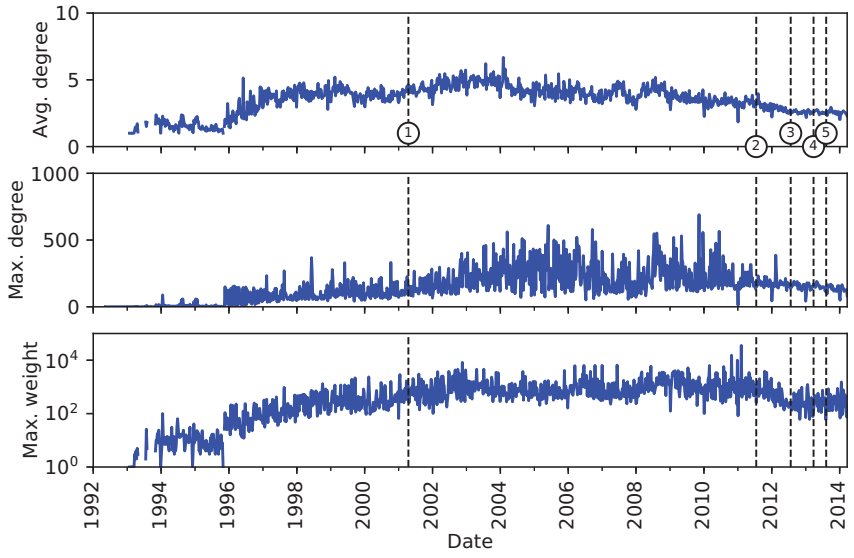

Figure 4.6: Time series of the avg. degree (top panel), max. degree (middle panel), and max. weight (bottom panel) for the bipartite graphs.

number of engineers in the one-mode projection graph. The degree of node i is $d_i^{\top}(k)$, i.e., its number of neighbors. This means the number of other engineers connected to i. The set of edges of of the graph $\mathcal{G}^{\top}(k)$ is $\mathcal{E}^{\top}(k)$. The number of edges is $|\mathcal{E}^{\top}(k)| = \sum_{i \in \mathcal{V}^{\top}(k)} d_i^{\top}(k)/2$. This means the total number of interactions in the one-mode projection graph. A connected component is a subgraph in which any two nodes are connected to each other by paths. This meas subgraphs made of engineers in which any two vertices are connected, i.e., engineers that work in related components. The average degree of graph $\mathcal{G}^{\top}(k)$ is $2 \times |\mathcal{E}^{\top}(k)|/|\mathcal{H}^{\top}(k)|$. This means the average number of activity of engineers in the graph. The maximum degree of graph $\mathcal{G}^{\top}(k)$ is the maximum number of neighbors in the graph, i.e., $\max\{d_i(k), \forall i \in \mathcal{H}^{\top}(k)\}$. This means the degree of the node with more interactions. The maximum weight of a graph $\mathcal{G}^{\top}(k)$ is the maximum weight of edges in the graph, i.e., $\max\{\omega_{ij}(k), \forall i, j \in \mathcal{H}^{\top}(k)\}$. This means the weight of the interaction with maximum strength.

Figure 4.7 (top) shows the observed number of nodes (i.e., engineers in the user graph). Figure 4.7 (middle) shows the number of unique edges representing the number of interactions between engineers. Figure 4.7 (bottom) shows the number of connected components in the one-mode projection graphs. In general, for the number of nodes and edges, there is an increase in these mea-

surements after roughly 2002. The tendency starts to decrease after approximately 2010 when other version control systems began to be adopted. Thus, after 2010, the data are a large sample rather than a comprehensive dataset. The movement of some core technologies to a different versioning system is reinforced by the continuous increase in the number of connected components in the one-mode projection graph, which indicates a less integrated core of components.

Similarly, Figure 4.8 shows the time series of average degree, maximum degree, and maximum weight respectively. Although there are several spikes for these measurements, we present an evaluation of the proposed algorithm, when these measurements inform the detection signature.

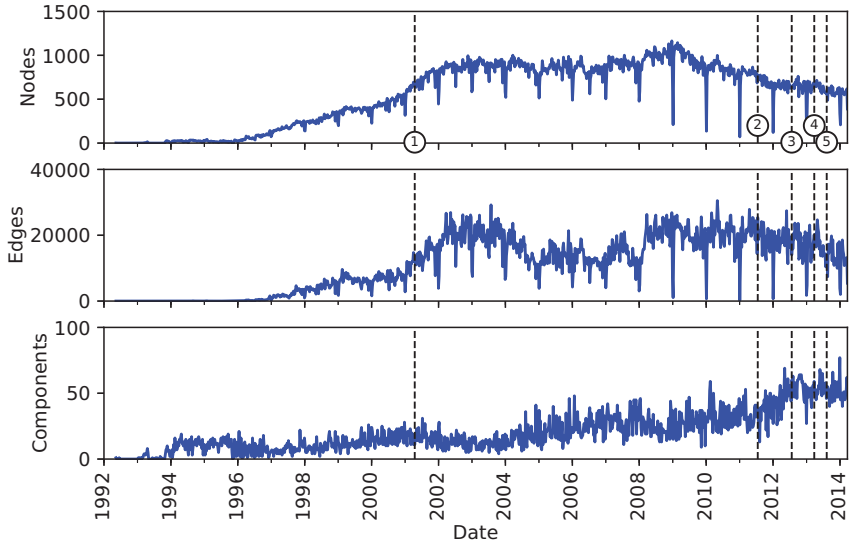

Figure 4.7: Time series of the number of nodes (top panel), edges (middle panel), and connected components (bottom panel) for the one-mode projection graphs.

4.4.3 Algorithm Evaluation

We applied the proposed algorithm for anomaly event detection by leveraging on the structural properties of the one-mode projection graphs. Note that one-mode projection graphs are derived from the original user-system interactions or bipartite graphs, so we compare the results of our proposed method with those properties that are measured in the equivalent one-mode projections. Our criteria for selection of anomalous time intervals is based on the idea of detecting observations that are far away from the median (for a specific time interval in which a one-mode projection

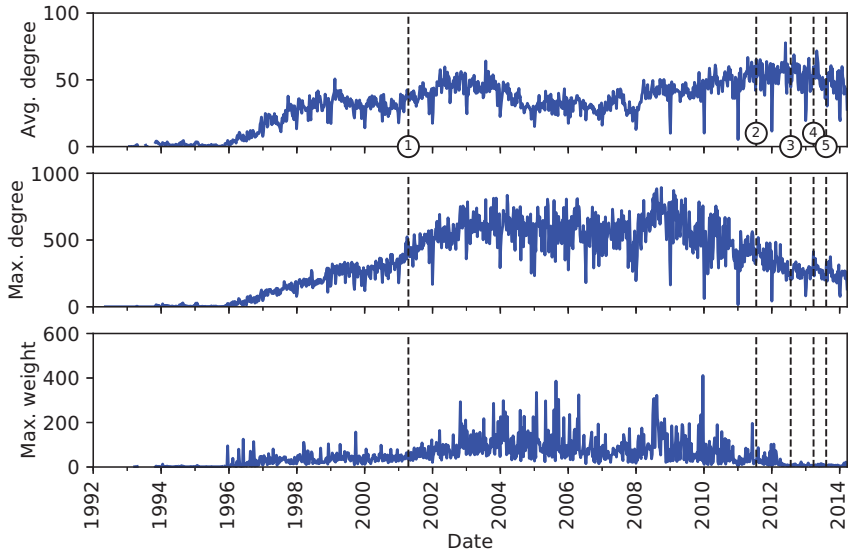

Figure 4.8: Time series of the avg. degree (top panel), max. degree (middle panel), and max. weight (bottom panel) for the one-mode projection graphs.

graph is generated) as we specify in Algorithm 4. Following similar visualization conventions that we used for the one mode projection graphs, in the following figures, the black horizontal line represents the median from the empirical observations. Each horizontal red band represents one standard deviation (the intensity of the bands is proportional to the distance with respect to the median). Remember that the standard deviation is estimated using the interquartile range of the distribution of these measurements. We estimated m_0, i.e., the length of the initial cumulative one-mode graph segment, by computing $\arg\max_{m_0} |C(\mathcal{G}_{\top}^{m_0})|$. That is achieved by the end of 2002, and it is the reason we report the following properties since January 1st, 2003.

Figure 4.9 shows the time series of nodes, edges, and connected components after the period of characterization of communities, i.e., the period of time comprehended between May 4, 1992 and December 31, 2002. As can be seen, even when there are some fluctuations in these measurements, the majority of the observations lay up to three standard deviations away from the median. This means that few time intervals were reported as anomalous during the observation period by relying in these properties.

We also performed similar experiments for the remaining graph-based properties, i.e., average degree, maximum degree, and maximum weight. In particular, Figure 4.10 summarizes these

findings. For both average degree and maximum degree, the algorithm did not report suspicious time intervals given that the signal does not exceed ± 3 standard deviations from the median. For the signal corresponding to the maximum weight, various spikes surprise the limits for detection.

Figure 4.11 shows the behavior for the proposed metric. Details on how this metric is derived are found in Equations 4.1 and 4.2. In particular, there are some spikes that exceed the threshold used by the algorithm and are close enough to the release of the precipitating events. These spikes suggest a drop in the number of edges between members of the same communities (conversely an increase in the number of edges between members of different communities) which, based on our proposal, means a diversified behavior, i.e., more interaction with different components.

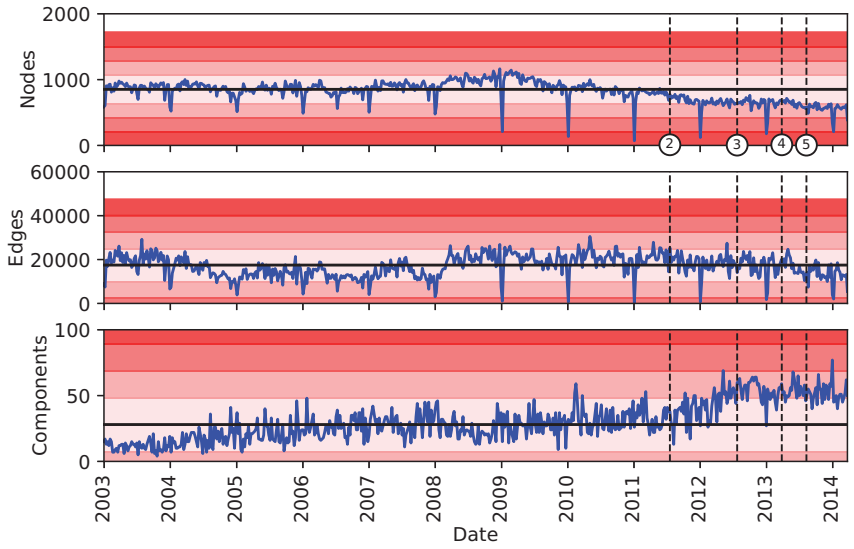

Figure 4.9: Time series of the number of nodes (top panel), edges (middle panel), and components for the one-mode projection graphs (bottom panel).

4.4.4 Algorithm Performance

We compare the performance of the proposed algorithm with the performance of a random algorithm. In particular, let the output of the random algorithm be $\hat{\mathrm{R}} = (\hat{\mathrm{R}}_1, \ldots, \hat{\mathrm{R}}_{\bar{\mathrm{n}}}) \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(0.5)$. This means that each time interval is equally likely to be selected as anomalous based on random chance.

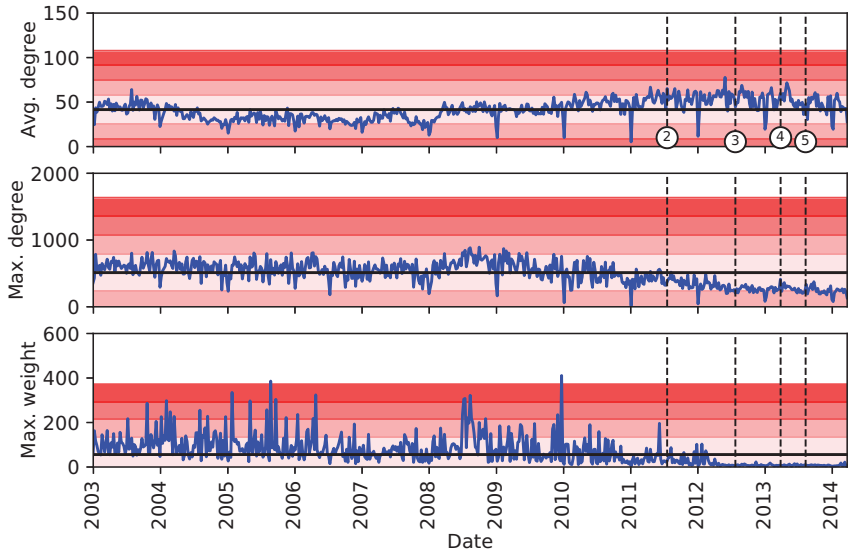

Figure 4.10: Time series of the avg. degree (top panel), max. degree (middle panel), and max. weight (bottom panel) for the one-mode projection graphs.

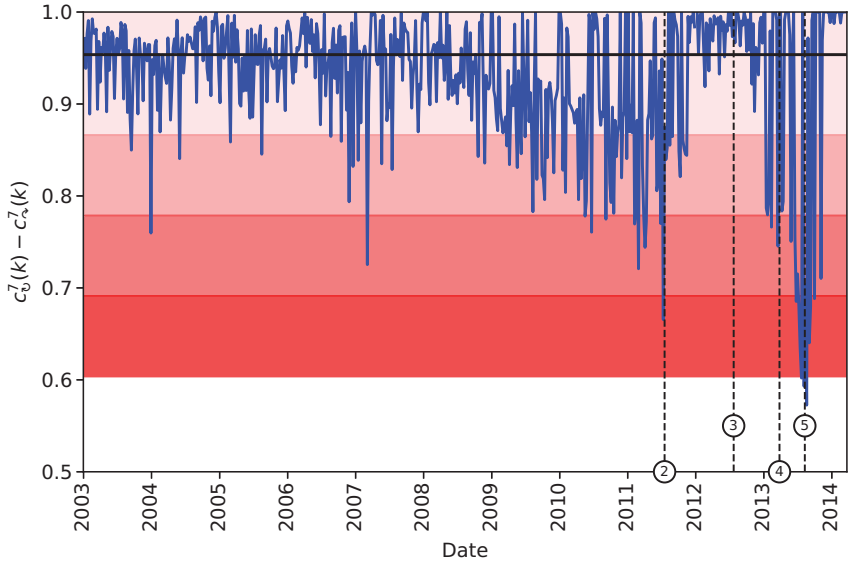

Figure 4.11: Time series of the intra- minus inter-community edge ratio for the one-mode projection graphs.

Performance for all the proposed algorithms is compared based on accuracy, precision, recall, and F1 score. These measurements were estimated using the TP, FP, FN, and TN derived from Algorithm 3.

Accuracy is the most basic measure of performance for classification. It quantifies the propor-

tion of correctly predicted positive and negative instances (i.e., time intervals classified as anomalous or not that were correctly classified). It is quantified as accuracy $= \frac{TP+TN}{TP+TN+FP+FN}$.

Precision quantifies the proportion of positive predictions that have been correctly classified. This means that if a considerable number of time intervals are erroneously classified as anomalous, then the algorithm has low precision. In other words, it is a measure of classification exactness. It is quantified as precision $= \frac{TP}{TP+FP}$.

Recall quantifies the proportion of actual anomalous intervals that have been predicted as positive. This means that if an insignificant number of time intervals are classified as anomalous but they are not, then the algorithm has low recall. In other words, it is a measure of classification completeness. It is quantified as recall $= \frac{TP}{TP+FN}$.

The F1 score conveys the balance between precision on and recall calculated through the harmonic mean. It is quantified as F1 score $= 2\frac{\text{precision} \times \text{recall}}{\text{precision}+\text{recall}}$.

Figures 4.12, 4.13, 4.14, 4.15 show the performance for different detection criteria, i.e., random, nodes, edges, connected components, average degree, maximum degree, maximum weight and the proposed approach under different detection resolutions. Performance in the random algorithm is calculated after $1,000$ realizations its evaluation. That means that for the random algorithm, we report on the mean and standard deviation on such measurements. As we might expect, the performance of the proposed approach starts increasing when the detection resolution is increased. For the maximum detection resolution that we used, i.e., 26m, the results of the proposed approach outperforms the other measurements with a F1-score of approximately 85.7%. Noticeably, the performance of the random algorithm is even higher than those based on graph measurements even when taking into account the effect of the standard deviations represented by the error lines.

Accuracy of detection methods based on the graph-based properties is high given that the majority of time intervals are not marked as anomalous based on the small number of precipitating events (which makes this an unbalanced dataset).

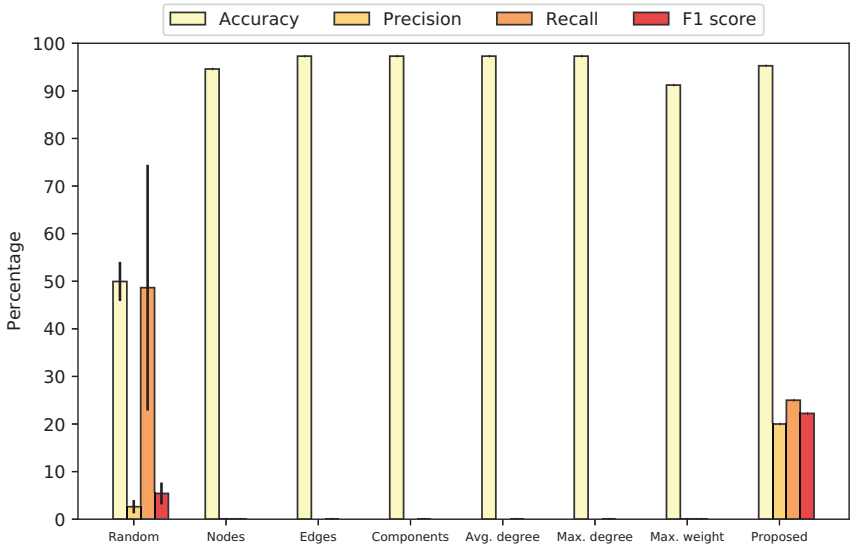

Figure 4.12: Algorithm performance for detection resolution 4m. Results of the proposed method are comparable with the ones of the random approach.

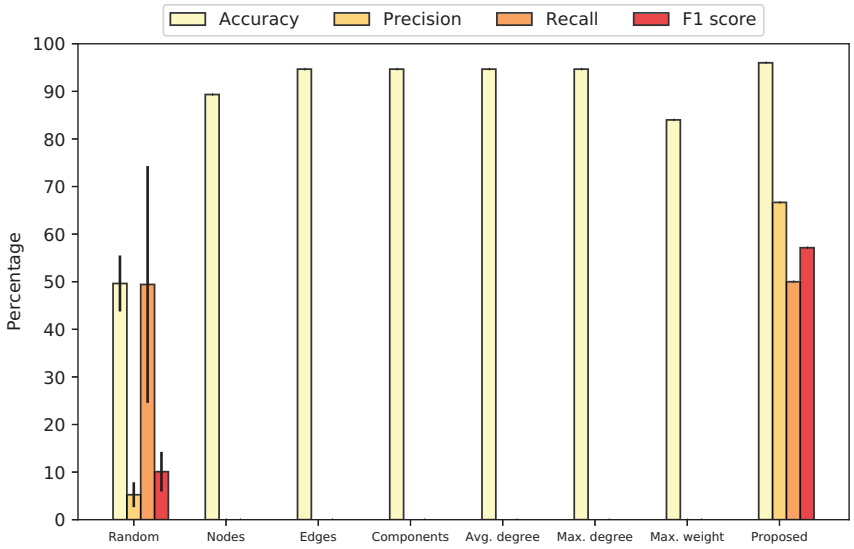

Figure 4.13: Algorithm performance for detection resolution 8m. Results of the proposed method begin being better than the ones of the random approach.

4.5 Conclusion

In this chapter, we have revisited the problem of insider threat event detection using graph mining analytics. To our knowledge, this is the first analysis proposing an insider threat detection method using temporal bipartite graphs to pinpoint malicious events. Our main contribution is the proposal

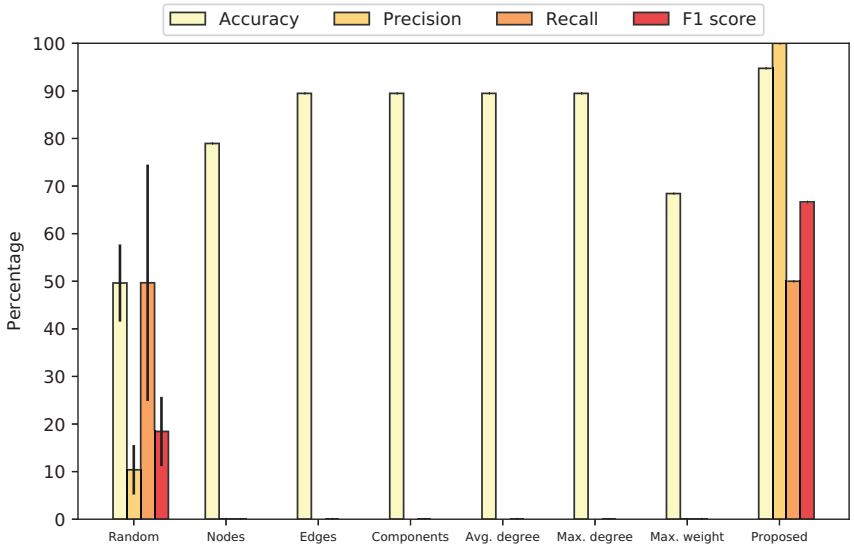

Figure 4.14: Algorithm performance for detection resolution 16m. Results of the proposed method begin being much better than the ones of the random approach.

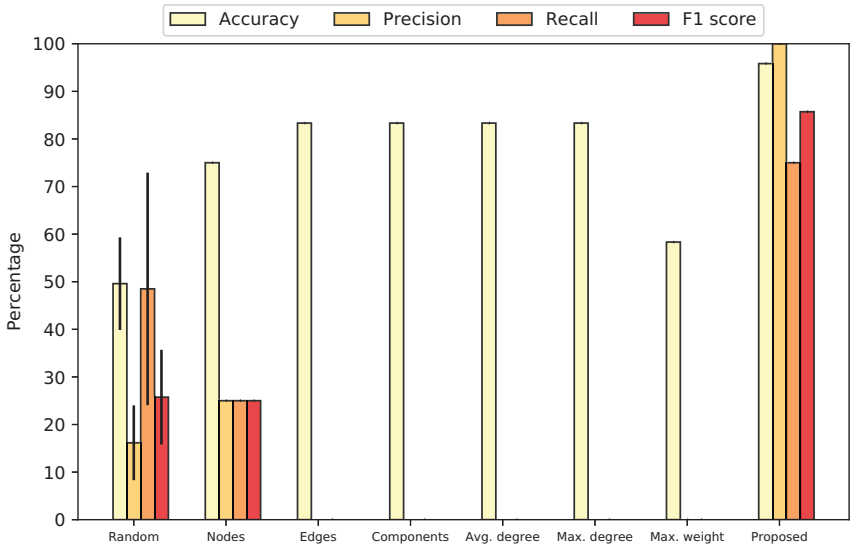

Figure 4.15: Algorithm performance for detection resolution 26m. We obtain about 92% of F1 score.

and evaluation of a generic analytical framework that builds on previous results in analysis of social networks to identify anomalous behavior by distinguishing access requests within and beyond a given community. We analyzed access to resources (i.e., code repositories) by employees (i.e., coders and engineers) using a time series of graph properties to pinpoint time intervals that identify suspicious insider behavior. The temporal analysis framework can be used with other datasets,

including by organizations with no interest in sharing internal logs.

One major challenge in identification of potentially malicious behavior is determining ground truth. Although catastrophic insider events are well documented, the regular exfiltration of data by insiders is less well documented. There is a dearth of data. To address this, we examined the incidences of suspicious activity and correlated these with events known to be correlated with increases in insider threat behaviors, specifically precipitating events. The decision criteria for identifying these time intervals is based on quantifying changes in the way in which employees interact with resources after precipitating events have been announced. This performance analysis framework can be used by any organization that has experienced precipitating events in order to test it for applicability to its own risks. Further, by altering the time period for the analysis, organizations can make their own trade-offs as to the level of activity that will result in investigation.

From our results, it is possible to see that the proposed framework is able to identify time intervals in which anomalous activity happens with a reasonable F1 score. We compare the performance of the proposed approach with anomaly detection approaches based on a naive random and edge density dependent statistics. Our approach outperforms these intuitive approaches giving us insights on the importance of the diversification of committing behavior on user-system interactions as a possible indicator of insider threat.

The main assumption behind the proposed approach is that insider threat events increase following certain types of events. Thus, counts of potentially malicious actions correlate with the announcement of precipitating events. Insider threat events are associated with stress or disgruntlement, and these responses are often triggered in the wake of precipitating events (e.g., [Mishra and Dhillon, 2006, Warkentin and Willison, 2009, Greitzer et al., 2009, Moore et al., 2013, Benjaminsen, 2017]). This abstraction allows us to test the hypothesis as to whether the diversification of the committing behavior of users changes after the presence of a precipitating event.

We have proposed a bipartite graph framework that learns regular community behavior based on the interactions of engineers and components, and analyzes the patterns of connections in and between communities. We then use this to examine a time period that includes major precipitating

events. As a result, the ground truth available for the analysis implemented here is the rate of insider risk in the organization after precipitating events. The validation of the model would be clear increases in the number of interactions across communities after precipitating events, and few increases without these.

Precision and recall together measure how often a threat is correctly identified and how often the non-malicious is correctly identified, i.e., no false positive or false negatives. This correctness is a significant challenge in detecting insider threats. Individual organizational tolerance for false positives versus false negatives may differ. Figures 4.12–4.15 show that this trade-off can be changed by altering the detection resolution for the analysis.

Our approach makes a well-grounded assumption about the overall rate of insider threats and examines aggregate detection after precipitating events. Alternative approaches use artificial data with anomalies generated based on scenarios and confidential data. Another alternative is using qualitative research and directly leveraging known cases. By definition, the artificial data and case studies can only address the insider threats that have been detected using other methods. A third approach examines private datasets which includes potential malicious insider behavior. Our results use a private dataset subject and temporal analysis to illustrate that insider behavior increases are correlated with what are known to be precipitating events.

Of the three methods to address suspicious insider behavior, reproducibility is a particular strength of artificial data and is a particular challenge to the third approach (i.e., the one used here). The challenge to the second (case studies) and third (confidential data) approaches are of reproduction and validation. To address these challenges, we will release the scripts used to implement this model on or before publication of this analysis. With the publication of our model as implemented, in addition to the description here, our analysis can be reproduced using any organization's private data. One goal in publishing this work is to encourage other researchers to use the model on the data available to them.

One requirement for this approach is adequate data to create the one-mode projection of the interactions between engineers from the bipartite graph of engineers and components. The current

dataset covers more than two decades of interactions with engineers and version control systems. The requirements for the minimal training dataset is an open question. With logging provided by version control systems, software organizations have adequate data. However, other organizations with different types of data may struggle to find the optimal input. Another question is the optimal size of a community or subgraph [Dong et al., 2012]. This is a parameter that will vary between organizations.

One possible weakness to this approach is that an organization with a systematic insider threat problem may be unable to use this as detection. Training for community detection requires the insider's behavior to be anomalous. For example, organizations with high levels of turnover may consistently see behavior that would be anomalous in another organization, one with has higher retention or a more careful workforce.

Our approach identifies behaviors as opposed to focusing on the motivation of an individual. As a result, the particular strength of this method is identification of a significant number of suspicious behaviors across the entire employee population. A weakness is that an employee who becomes slowly malicious and increases suspicious behaviors over time may be able to train the model of that organization not to recognize his behavior as anomalous. This attack would be mitigated by the characterization of others in organization (who cannot be controlled by the insider). As with all insider threat detection systems, any employee who has access sufficient to manipulate the input and output of the model itself can defeat the analysis.

It might also be the case that our assumptions are incorrect. It may be the case also that precipitating events are not the only triggers to this type of activity. If insider threats are a result or response to specific events, other specific events including employee dismissal, dispute with employers, perceived injustices, family problems, coercion, or new opportunities—as has been highlighted in [Nurse et al., 2014]—should be considered when evaluating the proposed approach.

5 Macroeconomics of Routing Anomalies[1]

"Yes, we have to divide up our time like that, between our politics and our equations. But to me our equations are far more important, for politics are only a matter of present concern. A mathematical equation stands forever."

— Albert Einstein

5.1 Introduction

In this chapter, we investigate why some countries are more likely to originate routing anomalies than others. To do so, we analyzed reported routing anomalies and macroeconomic indicators over a four-year period. There are well-documented hijacks resulting from errors, for profit, or for national security and national intelligence purposes. Any individual hijack could be an accident, a crime, or an attack. We report on an empirical investigation into the macroeconomics of routing anomalies that addresses these three explanations. If BGP anomalies are a result of limited technical competence, then countries with low levels of education and expertise may be over-represented. If BGP anomalies are a crime, used by criminals for profit, then economic theories and analytical approaches from criminology should be able to provide at least partial answers to this question. Alternatively, if BGP hijacks are primarily used by national intelligence agencies to attack either internal dissidents or other countries, then the presence of conflict would be an indicator.

Our analysis built on previous macroeconomic work in online crime using the factors that have been found to be significant in predicting malware, spam, and crowdsourced criminal labor (e.g. CAPTCHA solving). For crime, our data analysis evaluated if variables associated with

[1]The content of this chapter was initially proposed as a conference abstract [Moriano et al., 2016] and then published as a journal article [Moriano et al., 2017c] in collaboration with Soumya Achar and L. Jean Camp. Pablo Moriano is the primary researcher on both works and made all the analysis and figures therein.

(i) routine activity theory (fixed broadband subscribers, secure Internet servers—those that use SSL/TLS protocol for encryption and decryption to protect data from unauthorized interception), (ii) economic deprivation theory (GDP per capita by PPP), or (iii) structural theory (governance and education) were significant over multiple years. We found that two of the three theories were significant. Specifically, secure Internet servers, and density of fixed broadband subscribers were significant, supporting routine activity theory. Governance indicators were significant, supporting a structural theory of crime. Education alone was not found to be significant. Our findings suggest that strength of guardianship and weakness of governance do correspond to initiation of routing anomalies, supporting the possibility that these are driven by crime. In addition to these findings from regression analysis, clustering indicates that there are two groups of countries, one of highly developed and one of less developed nations. Cluster analysis showed some indication that civil war and surveillance may be correlated with hijacks. So while we cannot reject the possibility that these hijacks are a function of lower technical expertise, we also found no evidence for that argument. We found empirical evidence that some hijacks may be for profit.

5.2 Problem

Routing security is a complex technical, social, and economic problem. A given route anomaly may be an accident, a for-profit crime, or an intelligence operation. Understanding the economics and political dynamics of this threat can inform effective defenses. Solutions that mitigate risk from one possible source should not exacerbate the other. For example, a lack of expertise as a source of anomalies cannot be dismissed as an explanation out of hand. If so, security solutions that increase the complexity of network operations may perversely increase rather than decrease anomalies. Alternatively, trust models that are optimized to address for-profit crime may increase opportunities for intelligence operations.

This chapter presents an empirical analysis to better understand the nature of routing anomalies. We began with three hypotheses: anomalies result from (1) a lack of technical expertise, from (2) criminal activity, or from (3) national intelligence operations. We determined the factors that

would indicate expertise, criminality, and active intelligence, then operationalized these with linear regression and unsupervised clustering. We then identified correlations to reject (or not) these three possible explanations for anomalies. Using empirical data, we could not reject the second or third hypothesis. Although we found evidence of correlation with the second and third, this is not, of course, causation.

In a nutshell, the nature and distribution of control-plane anomalies are rife with open questions, and some of these are political and economic questions which can inform the design of security solutions. Here we address the question of the role of ecrime and politics in BGP anomalies using macroeconomic analysis and unsupervised learning.

5.3 Methods

We performed a longitudinal cross-country empirical analysis of the macroeconomic factors that correlate with the number of IP prefixes associated with potentially malicious control-plane events. That is, we focused on the originators and not the targets of BGP control-plane incidents.

The research presented here is based on 322,466 incidents that were identified not only as anomalies but also as possible hijacks by Argus [Shi et al., 2012]. This data may include false positives. A false positive, in this case, is the identification of a routing event as anomalous when it is not. Possible sources of false positives include route flapping, organizational changes, or unusual but not malicious responses to changes in the control-plane from outages or congestion. Our investigation is based on the argument that the possibility of BGP attacks systematically being used for profit and for surveillance cannot be dismissed. BGP control-plane attacks as a tool for ecrime or espionage has implications for investments, policy, possible targets, and possible solutions.

We used macroeconomic variables that have been found to be relevant to global analyses of ecrime, grounded in traditional theories of crime [Pratt and Cullen, 2005], as well as prior research focused on understanding the economic incentives behind malware [Garg et al., 2013]. We then used standard regression techniques, specifically multiple linear regression using ordinary least

squares, to examine these theoretically important independent variables. As a dependent variable, we considered the number of hijacked IP prefixes that originated in a particular country using the data set provided by Argus. We analyzed data over a nearly four-year period, from 2011 to 2014.

5.3.1 Data Sources

5.3.1.1 Independent Variables

The independent variables used in this research are based in traditional theories of crime, in particular: (i) routine activity theory, (ii) economic deprivation, and (iii) social support (also known as altruism theory). These variables have been instrumented using publicly available data from the World Bank.[2] Although there is a dispute over the way some of these measures are implemented, biases in these measures are likely consistent across countries and over time.

The World Bank provides a consistent and systematic measure of country-level macroeconomic variables, and has been used widely in economic, criminology, and jurisprudence research [Ika et al., 2012]. Equally important, the World Bank provides uniform and consistently available measures that are likely to be reproduced by other researchers. We grouped the independent variables into factors and described them in detail below (see Table 5.1).

The routine activity theory of crime states that the probability of a crime is a function of motivated offenders, available targets, and lack of guardianship [Felson and Cohen, 1980]. The first independent factor is **availability** of users, i.e., a larger population of Internet users is more likely to produce more hijack attacks. This means that we make no assumptions about the distribution of motivation between categories or classes of users but rather assume that with more users there is a larger pool of people who could be motivated to commit online crime. To determine if this is a significant variable, we considered the number of fixed broadband Internet subscribers (FBIS) in a given country. We normalize the number for comparisons across countries by evaluating the number of fixed Internet broadband subscribers per 100,000 people (FBISper100).

[2]Available for public download at `http://data.worldbank.org/`

A second independent factor is **guardianship**. Guardianship has been modeled in other works as community oversight. High guardianship indicates that the community has and enforces norms of non-criminal and safe behaviors. In the case of online crime, this can be indicated by private investments in the existing information and communications technology (ICT) infrastructure. We operationalized this factor using the measure of secure Internet servers (SIS) and SIS per million people (SISperM) to account for the differences in population. The World Bank defines SIS as "servers using encrypting technology in Internet transactions."

Economic deprivation theory suggests that crime is driven by blocked legitimate opportunities or **economic** deprivation [Blau and Blau, 1982]. This is the third independent factor in our analysis, and the only factor we examine for the economic deprivation theory of crime. Certainly, a lack of economic opportunity encourages individuals to perform criminal activities if these are profitable. To measure this factor, we consider the gross domestic product (GDP) per capita by purchasing power parity (PPP). Simultaneously, GDP per capita by PPP (GDP PPP) indicates the relative deprivation with respect to other countries given the low cost of global connectivity.

The social support or altruism theory posits that an individual lack of economic resources can be alleviated by appropriate **governance** through public investments [Cullen, 1994]. This motivates the inclusion of a fourth independent factor in our analysis. For ecrime, government influences take the form of subsidies or incentives for the adoption of technologies. Conceptually, when a governance environment encourages the adoption of technologies, it also encourages increased access to ICT technologies. Under the altruism theory of crime, this would also decrease the incentive for routing attacks. To capture this factor, we operationalized government support using a subset of World Governance Indicators (WGI) [Kaufmann et al., 2011]. The subset we included consists of (1) government effectiveness, (2) regulatory quality, (3) rule of law, and (4) control of corruption, i.e., perception of corruption within a country. Government effectiveness captures the perceived quality of the services, policy formulation, and the credibility of the governance. Regulatory quality refers to the perceived degree of alignment toward the development of the private sector. Rule of law measures the degree to which the legal framework is operationalized in prac-

tice. For example, corruption can decrease the effectiveness of the legal framework. Control of corruption measures the misuse of public power for private gain.

Finally, recall that one possible explanation of routing anomalies is a lack of **technical competence**. This is the fifth and final independent factor. Superior governance in the specific domain of ICTs would be indicated by high levels of local expertise. Local expertise and ability in computing is commonly identified in macroeconomic analysis by higher levels of exports in those industries. To measure local skills, we used the percentage of exports of computers, communications, and other services (CCS) as the corresponding macroeconomic variable. This measure is intended to quantify the degree to which a country has developed ICT markets. A larger ICT market implies more personnel who are trained in basic security practices. In fact, hijacking routes requires a minimum level of expertise in security and minimum technological requirements to perform, particularly in comparison with simple email confidence scams such as advance fee fraud. In contrast, creating a route leak requires a lack of expertise or due care. We recognize that guardianship and competence interact, and address this in our analysis.

Table 5.1: Five-dimensional regression model variables.

Model factor	**Variable name**	**Acronym**
Availability (AVA)	Fixed broadband Internet subscribers	FBIS
	Fixed broadband Internet subscribers (per 100 people)	FBISper100
Security of ICT infrastructure (SEC)	Secure Internet servers	SIS
	Secure Internet servers (per one million)	SISperM
Economic resources or affordability (ECO)	GDP per capita by PPP	GDP PPP
Governance of legal framework (LEG)	World Governance Indicators	WGI
Security skills or education (EDU)	Computer, comm., and other services (% exports)	CCS

5.3.1.2 Dependent Variable

The dependent variable corresponds to the number of unique affected IPv4 prefixes by country (based on the origin of the malicious prefix), i.e., each event corresponds to a prefix being rerouted.

We depended upon the Argus data set for the identification of anomalies. For the purpose of collecting routing anomalous events, we compiled data over the 43 months from June 2011 to December 2014 from the Argus[3] API, as of December 20, 2015 [Shi et al., 2012, Xiang et al., 2011].

Argus defines three categories of anomalies. Specifically, the data we used relies on the output of the Anomaly Monitoring Module (AMM). The AMM detects (1) origin, (2) adjacency, and (3) policy anomalies. An origin anomaly is detected when an AS is advertising a prefix that it does not own. To identify origin anomalies, Argus maintains a database that tracks the expected origin of prefixes. An adjacency anomaly occurs when there is a path which has two AS numbers that are not normally seen to be connected. This could be considered a bad or an unlikely hop. To detect adjacency anomalies, the AMM monitors changes in AS-path segments. Given the exponential number of possible combinations of segments in the updates, the AMM only verifies neighboring pairs in an AS-path. To detect policy anomalies, the AMM takes into account the type of commercial relationship between ASes. Specifically, customers are not expected to announce routes from their providers. A bad segment of three hops means a policy anomaly. Note that by considering these three types of anomalies, the AMM is able to detect a wide range of hijacks and other events.

Based both on Argus' description of their practices and the sheer number of events reported, we can assume flapping is filtered out. (Flapping occurs when a route oscillates between two or more possible paths that are themselves relatively stable.) We do acknowledge that some BGP alarms might be due to benign BGP engineering practice or misconfiguration, including events specifically classified as anomalous in this data. The common characteristic of the incidents examined below is that the IP prefixes and ASes are used to divert Internet traffic.

The hijacks reported by Argus are based on a mixed strategy between control- and data-plane methods, i.e., it correlates control-plane anomalies and data reachability to improve detection accuracy. This strategy enables Argus to distinguish hijacks from other types of anomalies. In the subsequent analysis, we did not consider the effects of the false positives detected by the Argus

[3]Available for public download at `http://argus.csnet1.cs.tsinghua.edu.cn/`

system. That is, we assumed that each anomaly identified by Argus was, in fact, identified correctly.

5.3.1.3 Data Preprocessing

In our work, we focused only on *origin anomalies*. Origin anomalies occur when there are multiple announcements of an IP prefix and Argus judges one of these as malicious. Thus, each event used in the data set corresponds to a unique prefix being rerouted. If there was an attack that resulted in multiple prefixes being rerouted, then that would appear in the data as two attacks.

Our data set excludes large-scale and easily-identifiable routing anomalous events to prevent these black swan events from biasing the results.[4] So, had we included the 179,000 prefixes announced by Malaysia [Toonk, 2015a], the Chinese announcement of millions of addresses [Hiran et al., 2013], route leaks [Toonk, 2012], or large-scale outages [Dainotti et al., 2011] this would have biased the results. Instead, our data set of "events" or "incidents" consists of those incidences identified by Argus as origin anomalies that were not large-scale events but focused on a single prefix. If the same prefix is announced again in a different timestamp, it was counted as a different event. In addition, the size of the prefix does not have any effect in the counting process, i.e, we did not distinguish between the announcement of a /18 from that of a /24 of event prefixes but treat them equally.

Our macroeconomic analyses required correlating an AS with a jurisdiction. The jurisdiction of the event is defined as the country from which the bogus route was announced. The mapping of IP addresses and ASes to country has been widely used in industry [Microsoft, 2011, Anti-Phishing Working Group (APWG), 2015, Cisco Systems, 2015] and academic macroeconomic research, e.g. [van Eeten et al., 2010]. Geolocation is another related active research area, and improvements in geolocation could enhance this analysis. Here our approach to location is grounded in concurrent

[4] In the context of this research, a large-scale routing anomalous event is an incident that compromised thousands of IP prefixes in a single announcement. It constitutes an example of a black swan. A black swan is a highly improbable and high-impact event, such as the Lehman Brothers collapse in the U.S. or Malaysia announcing 179,000 prefixes.

industry practice and related research [Quan et al., 2014].

We inferred the mapping between ASes and country membership from a data set provided by the Cooperative Association for Internet Data Analysis (CAIDA) [CAIDA, 2015]. The geolocation data was also investigated over the fall of 2015 with multiple downloads of regional Internet registry (RIR) data. One goal in implementing downloads was to avoid the "phony, yet plausible, AS origins" that are used for ecrime. A false claim to an IP prefix would have to remain stable for a long time to pollute our data. There also would have to be a significant number of these to bias the results of our analysis.

Most of these anomalies in the Argus database are short-lived, on the matter of seconds. Previous anecdotal discussions [Madory, 2015a] described cases that appear to be both intelligence activities and criminal activities. Like the Bitcoin miner attack, apparently criminal incidents are short-lived. Other than anecdotes, we have no basis for asserting that a short-lived hijack is a crime while a long-lived hijack is evidence of international intelligence activity. Given that there is clearly malicious activity that occurs in short bursts, we included all the data in our analysis rather than creating an arbitrary threshold.

5.3.2 Statistical Model

Equation 5.1 summarizes the behavior that we captured.

$$\begin{aligned} A &= \beta_0 + \beta_1 \times \text{AVA} + \beta_2 \times \text{SEC} + \beta_3 \times \text{ECO} \\ &+ \beta_4 \times \text{LEG} + \beta_5 \times \text{EDU} + \epsilon \end{aligned} \tag{5.1}$$

In Equation 5.1, A refers to the number of hijacked prefixes (i.e., the dependent variable). Every independent variable is captured by the factors AVA, SEC, ECO, LEG, and EDU respectively as we defined in Table 5.1. The terms $\beta_0, \ldots, \beta_5$ are the regression coefficients of the model. In addition, ϵ is the error term. Equation 5.1 was evaluated using multiple linear regression with ordinary least squares (OLS). In an effort to validate the OLS assumptions, we examined the model for the

(absence) of multicollinearity[5] and heteroskedasticity[6].

To address the absence of multicollinearity in the predictors, we calculated the variance inflation factor (VIF) per year and in the aggregate data. We found that the four indicators of WGI are highly correlated, i.e., government effectiveness, regulatory quality, rule of law, and control of corruption have VIF > 5 [Zuur et al., 2010]. Thus, in the subsequent analysis, we combined all of these indicators into a single score by adding them up. We called this new variable WGI. For the remaining predictors, the VIF values did not indicate that multicollinearity is an issue, i.e., FBIS, FBISper100, SIS, SISperM, GDP PPP and CCS have VIF < 5.

We examined the distribution of the residuals of the model to check the absence of heteroskedasticity. We tested for normality using the Shapiro-Wilk test under the null hypothesis that the residuals were normally distributed. For the model described by Equation 5.1, the test for the distribution of the residuals produced a p-value of 0.464. This suggests that we failed to reject the null hypothesis that residuals are normally distributed. In other words, there is not enough evidence in favor of heteroskedastic errors, and the condition is satisfied. To avoid the regression being affected by particular outliers, we log transformed independent variables with a long-tail distribution. We determined these variables by examining their distributions in box plots and identifying outliers. The independent variables that were log transformed were (1) GDP PPP, (2) FBIS, (3) SIS, and (4) SISperM.

Similarly, given that the OLS regression makes additional assumptions, we explored the distribution of the dependent variable. In particular, OLS assumes that the dependent variable is continuous and normally distributed. In doing so, we tested for normality using the Shapiro-Wilk test per year and in the aggregate. We found that the normality assumption was not satisfied, in other words, the p-value ≈ 0. This indicated that the dependent variable is unlikely to be nor-

[5]Multicollinearity is observed when two or more independent variables are highly correlated, i.e., at least one of the independent variables can be expressed as a linear combination of the rest to a statistically significant degree. For collinear independent variables, the results of the OLS regression may be misleading.

[6]Heteroskedasticity is observed when the residuals (errors) of a model are not normally distributed. A test of absence of heteroskedasticity determines the regression model's ability to predict values of a dependent variable over all its range. For heteroscedastic residuals regression results should not be trusted.

mally distributed. Thus, the dependent variable was log transformed to satisfy the assumptions underlying the linear regression.

5.4 Results

5.4.1 Anomaly Time Series

The total number of IP prefixes that were impinged by an anomaly for each year during the observation period is 54,087, 90,607, 105,600 and 72,172 for 2011, 2012, 2013, and 2014, respectively. The data have a surprising level of variance. In particular, there is an increase of 67.5% between 2011 and 2012, an increase of 16.5% between 2012 and 2013, and a decrease of 31.7% between 2013 and 2014. This suggests that although the number of affected prefixes has been growing since 2011, it suddenly dropped at some point in 2014.

To further investigate the decrease in the number of hijacks in 2014, we analyzed the daily time series of anomalies. Fig. 5.1 shows the number of reported events during the observation period, i.e., the number of hijacked IP prefixes. The dashed line represents the unweighted LOESS fit. The shaded area corresponds to the 95% confidence interval of the regression model. It is worth noting that the data have very high variance. In particular, the number of hijacks spans four orders of magnitude. This means that reporting the average number of events is highly influenced by outliers and can be misleading. To take into account this constraint, we applied a non-parametric regression method—unweighted LOESS—to perform local regression and derive a non-linear fitted smooth curve (the dashed one) that captures the trends in the time series. We also plot the 95% confidence interval of the regression line in the shaded area. Although the smooth curve does not reveal a significant trend, it is possible to infer that in 2014, there were two periods in which there are not records of incidents. These periods correspond to the days between January 3 to April 22 and May 22 to July 29. We iterated the analysis to confirm this empirically. The decrease in the number of incidents in 2014 is likely a result of some downtime because there appears to simply be missing data. Only Argus could confirm this; however, it offers one possible explanation. We cannot reject

the explanation that hijacks have become more difficult to detect, but we find it unlikely. The rate of incidents appears to be increasing over time; however, missing data makes this an uncertain observation.

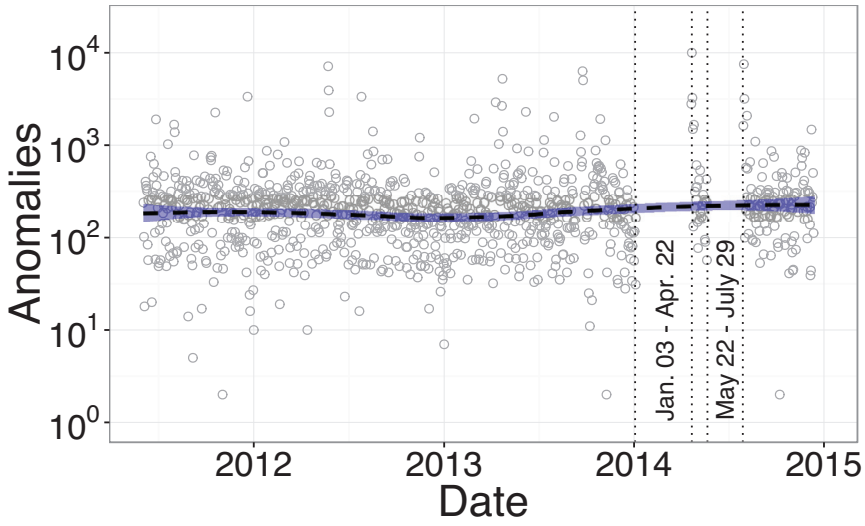

Figure 5.1: Daily number of anomalies. The dashed line represents the unweighted LOESS fit. The blue band corresponds to the 95% confidence interval of the unweighted LOESS fit.

5.4.2 Anomaly Distribution

To take into account the origin of the bad routes, we mapped the AS origin of the bad route to the country corresponding to the one in which the AS is registered (based on a CAIDA data set) [CAIDA, 2015]. We then measured the corresponding number of incidents that were reported in any given country. Fig. 5.2 shows the Complementary Cumulative Distribution Function (CCDF)[7] of the number of incidents per country during the observation period. The distribution is heavy-tailed suggesting that the majority of the reported hijack incidents originated from a small set of countries.

We determined the distribution that best fit the data by estimating parameters for different possible distributions and evaluating the goodness of fit. Specifically, we examined the best fit

[7]For a random variable X, the CCDF is the probability that X will take a value greater than a fixed value x, i.e., $\Pr(X > x)$.

parameters for the empirical distribution by using the maximum likelihood among a set of power-law, log-normal, and exponential parametric distribution candidates. They are very well known to be a good fit for heavy-tailed distributions. We followed the procedure explained in [Clauset et al., 2009]. Second, we calculated the goodness-of-fit for the proposed distributions by using the Kolmogorov-Smirnov (KS) test [Massey, 1951]. The KS test evaluates the hypothesis that two samples of data are drawn from the same distribution. In this case, we tested if the observed distribution was statistically significantly different from the best fit of the parametric model. In particular, we computed the proportion of times that the test failed to reject the hypothesis that the distribution of compiled data and the synthetic data are drawn from the same distribution for a significance level of 0.05. In tests of 10^4 synthetic distributions, the analysis of the synthetic data under the power law model failed to reject the null hypothesis in 99% of the cases, followed by log-normal at 98%, and exponential at 0%. From this analysis, we ruled out the exponential distribution as a potential candidate.

The dashed line in Fig. 5.2 shows the CCDF of the best power law model for the aggregate date, including all observations between 2011 and 2014. The estimated parameters of the power law distribution are $\alpha = 2.13$ for the scaling exponent and $x_{\min} = 958$ for the threshold at which the power law begins.[8] Although the best fit is given by the power law distribution followed by a log-normal distribution, there was not enough evidence to entirely rule out the option of the log-normal distribution. Moreover, there are still some deviations from the dashed line in the tail, i.e., for countries with a large number of hijacks.

To determine how these distributions changed over time (and the possible transition from one distribution to another), we applied the KS test to determine if the distribution of hijacks per country changed from one year to the next through 2011 to 2014. We found that there were no comparisons where we were able to reject the hypothesis that the distributions were the same (p-value $\gg$ 0.05). Fig. 5.2 illustrates this fact by showing that the CCDF over the period of study is visually

[8]Remember that a random variable X that follows a power law distribution has a probability density function given by $\Pr(X = x) = x^{-\alpha}$.

similar.

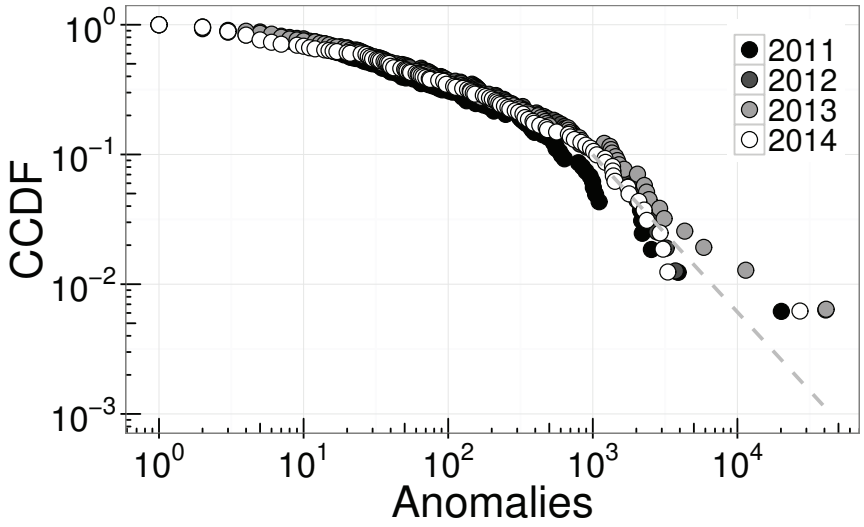

Figure 5.2: CCDF of anomalies originated per country.

To further study the behavior of the countries that are the origin of the majority of the anomalies, we plot the number of anomalies per country for countries that are in the upper 2.5% tail of the anomaly distribution during a particular year. Specifically, Fig. 5.3 shows the changes for the countries that are the origin of the majority of the anomalies. This shows the number of reported anomalies per year for those countries that are in the upper 2.5% tail of the anomaly distribution for a specific year. Interestingly, the U.S. is the country to which the majority of the anomalies are attributed throughout the observation period. Brazil and India appear in three of the four years during the observation period, although their relative ranks change. Russia and Turkey appear twice, and China and Romania each show up once.

5.4.3 Regression Analysis

We ran a multiple linear regression for the aggregate data during the observation period using the entire set of measures from 2011 to 2014. Remember that the target variable is the number of hijacked prefixes in a particular country i.e., OLS was applied to Equation 5.1. Table 5.2 shows the OLS regression estimates. The model estimates and standard errors have been presented by taking

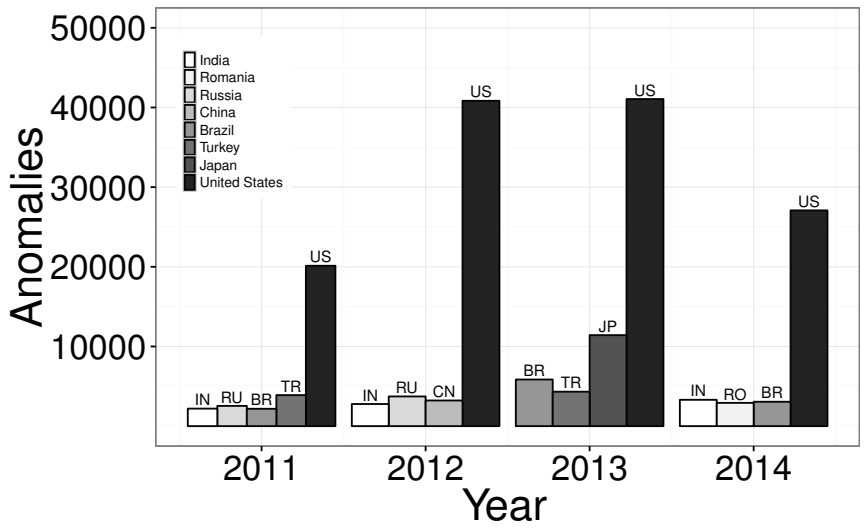

Figure 5.3: Anomalies per year for countries that are in the upper 2.5% tail of the anomaly distribution.

heteroskedasticity into account. The linear regression quantifies the individual relative importance of a single feature when considering the effect of the others constant.

There are three statistically significant factors, nominally, SIS, WGI, and FBIS, in order of decreasing importance. We did not count as a separate predictor SISperM because this is a normalized measure of SIS (an indirect measure). The coefficients of the OLS suggest that although there is a positive association between SIS and FBIS with the number of hijacked prefixes (i.e., countries with higher SIS and FBIS are more likely to produce more incidents), the association with WGI was negative (i.e., countries with poor governmental practices are more likely to produce more incidents). This analysis also suggests that our hypothesized macroeconomic factors explained a significant amount of variance in the number of incidents originating in different countries (approximately 75.3% with p-value ≈ 0).

We further explored the relationship between the number of anomalies and the number of SIS. Fig. 5.4 shows that there is a power relationship between the two variables (given by the linear tendency in a log-log plot) for both the aggregate data and for each of the individual years. For the aggregate, the association between the two variables (in the log scale) has a Pearson correlation coefficient of 0.79, which confirmed the idea of a strong association between the two variables.

Table 5.2: OLS regression estimates.

Variable	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.7231	0.7913	-2.18	0.0300*
log(FBIS)	0.1700	0.0545	3.12	0.0019**
FBISper100	-0.0052	0.0096	-0.54	0.5882
log(SIS)	0.7561	0.0654	11.57	$\approx 0^{***}$
log(SISperM)	-0.1758	0.0677	-2.60	0.0098**
log(GDP PPP)	-0.0090	0.1145	-0.08	0.9372
WGI	-0.1222	0.0326	-3.75	0.0002***
CCS	0.0035	0.0024	1.47	0.1429

Signif. codes: ***p < 0.001, **p < 0.01, *p < 0.05

Residual standard error: 1045 on 408 degrees of freedom

Multiple R-squared: 0.7576, Adjusted R-squared: 0.7534

F-statistic: 182.1 on 7 and 408 DF, p-value: < 2.2e-16

Fig. 5.4 also reinforces the results of the regression analysis. In other words, it suggests that countries with a high number of SIS are also the countries with a high reported number of anomalies. In this plot, we highlighted the country with the highest density of secure servers and the larger number of produced anomalies during the observation period, which was the U.S. every year.

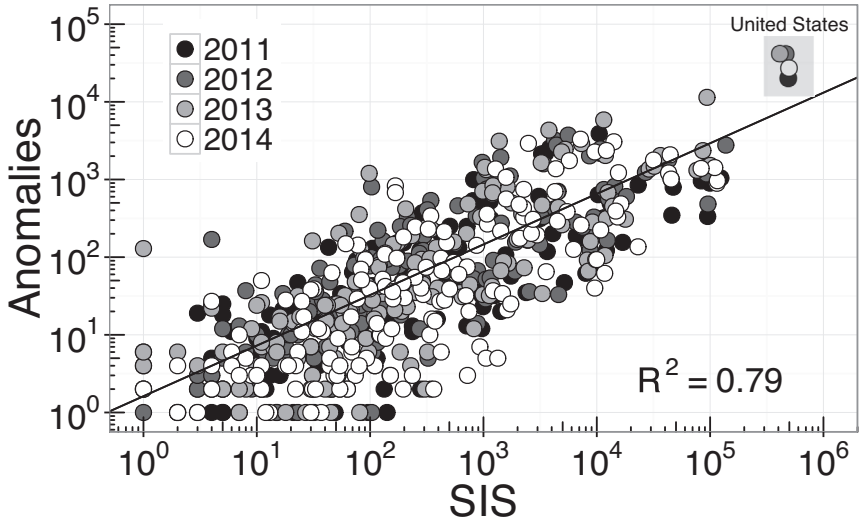

Figure 5.4: Anomalies vs. SIS per year per country.

The top five countries in terms of the number of SIS during the observation period remained the same: U.S., U.K., South Korea, Japan, and Germany. The U.S. reported half a million, with the others reporting on the order of hundreds of thousands. As we might expect from the positive

association revealed in Fig. 5.4, the U.S. is the country with the highest number of SIS and the highest number of anomalies (see Fig. 5.3). Japan is the other country that reveals a similar pattern with respect to a positive association between the number of SIS and control-plane anomalies. On the other hand, Germany, South Korea, and the United Kingdom report a high number of SIS, but here this does not correspond to a larger number of anomalies (see Fig. 5.5).

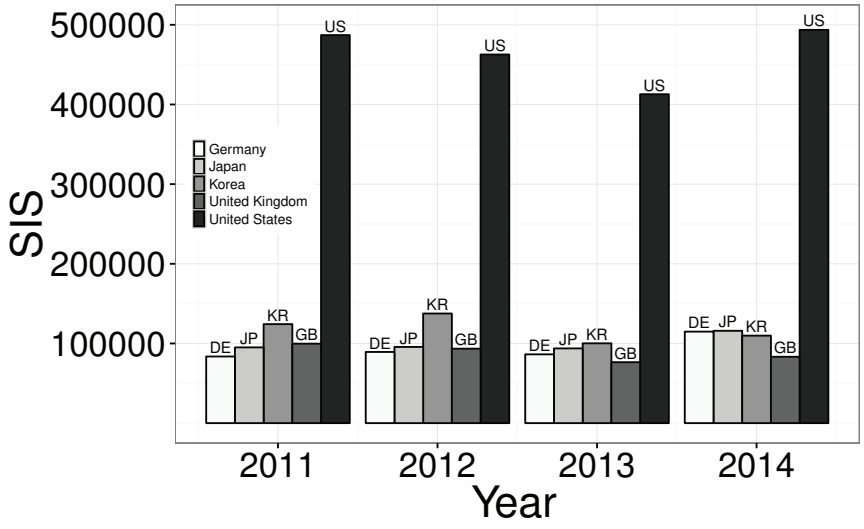

Figure 5.5: Distribution of the number of SIS per year for countries that are in the upper 2.5% tail of the SIS distribution.

We then investigated the association between the number of SIS and WGI. Fig. 5.6 shows the relationship between these two variables for every country during the observation period. Note that we only considered the number of SIS and WGI because they were the most significant predictors in the regression, i.e., we did not explore the effect of FBIS. Although there is a positive association between SIS and the number of hijacked prefixes, and a negative one between WGI and the number of hijacked prefixes (see Table 5.2), the relationship between these two independent variables is exponential in the WGI (given by the linear tendency in the log-linear plot) both for the aggregate data and for each of the individual years. The correlation for the aggregated data has a Pearson correlation coefficient of 0.68. This fact reinforces the idea of a strong association between the two variables. It also suggests that, given the data, it is more likely that countries with poor indices of governance also have a smaller number of SIS and vice versa. But more importantly, any change

in the WGI of a country is exponentially amplified and reflected in the number of SIS. It appears that small changes in the perception of governmental practices result in significant changes in the generation of anomalies, and thus significant changes in creation of risk in the global control-plane.

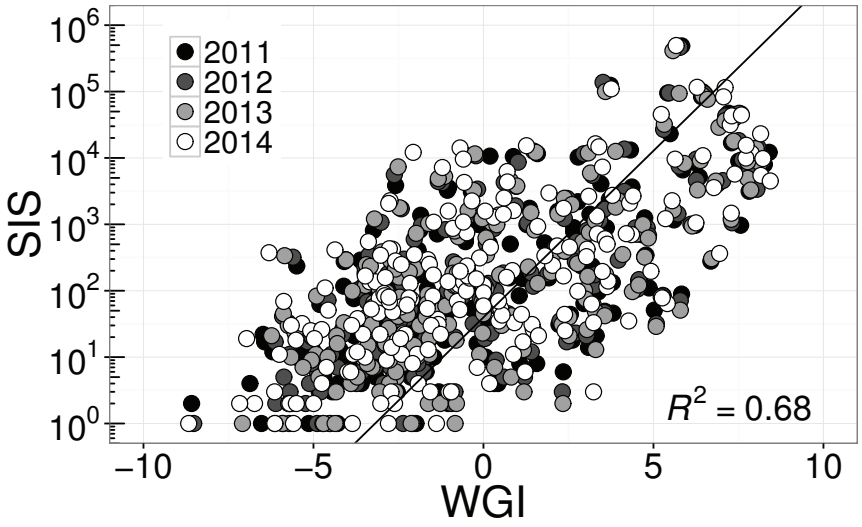

Figure 5.6: SIS vs. WGI for all countries over all years.

5.4.4 Cluster Analysis

For the cluster analysis, we first normalized the anomalies per country by the total number of ASes that belong to that country during a certain year. This transformation allowed us to have a fair metric to evaluate the likelihood of a country being the origin of the anomalies based on their technological resources. Fig. 5.7 illustrates the differences in the anomaly/ASes ratio across the SIS and WGI plane for 2011. A higher intensity (i.e., darkness) in the data points represents a higher anomaly/ASes ratio at the country level. The plot also depicts the contour of the two-dimensional density estimation of the distribution of the data points. For those unfamiliar with unsupervised learning, a description of the use of unsupervised learning to categorize data can be found in [Hastie et al., 2016, Chapter 14].

We did not implement an aggregate cluster analysis for the entire period because of the apparently missing data as shown in Fig. 5.1. We have implemented cluster analysis for 2012 and 2013

finding the same clusters and same outliers despite the changes in rank shown in Fig. 5.4. The only difference is that Somalia was not an outlier in 2013. All graphs were remarkable for their similarity. The following discussion of cluster analysis applies to 2011-2013, with data missing in 2014, and 2015 subject to further analysis.

Fig. 5.7 shows that there are two main clusters based on this distribution. The first cluster corresponds to countries with high WGI and more than 10^3 SIS, i.e., the first quadrant. Not surprisingly, the majority of the countries in this cluster are developed countries. The U.S. and South Korea are outliers of this cluster and are labeled in the plot. Although the U.S. has a better WGI indicator and a larger number of SIS than South Korea; South Korea has a higher ratio of anomaly/ASes (which is represented by a darker point). Recall the U.S. and South Korea are outliers in all our cluster analyses. South Korea similarly has a higher ratio of hijack/ASes for all years.

The second cluster represents mostly countries with low WGI and less than 10^3 SIS, i.e., the third quadrant. The majority of countries in this cluster are developing countries. Comoros has been labeled in the plot with the highest ratio of anomaly/ASes during 2011 despite democratic elections of presidents who rejected violent extremism and the first peaceful transfer of power in 2010. In 2011, Comoros saw the return of the United Nations, the Peace Corps after 20 years, and such basic governance activities as setting up the first national parks. Terrorism warnings for travel in Comoros are low, and there are rare incidents of mass violence. Thus, this may be an anomaly, a result of low Internet penetration, or an echo of the irregularities of the 2010 election. Alternatively, and the argument the authors find most compelling is that Comoros is a geographical anomaly. The island is a nexus of global fiber traffic despite a small number of ASes. For Comoros, the sheer density of connectivity has no correspondence to number of ASes.

Somalia is an outlier of this cluster in 2011 and 2012. Somalia was not an outlier in 2013 despite a continued lack of effective central government, prevalence of violent extremists, and a civil dispute over the state of Somaliland. It may be worth noting that Somalians have recently had some success in leveraging ecrime, however, their technical expertise as observed is quite limited [Gallagher, 2016].

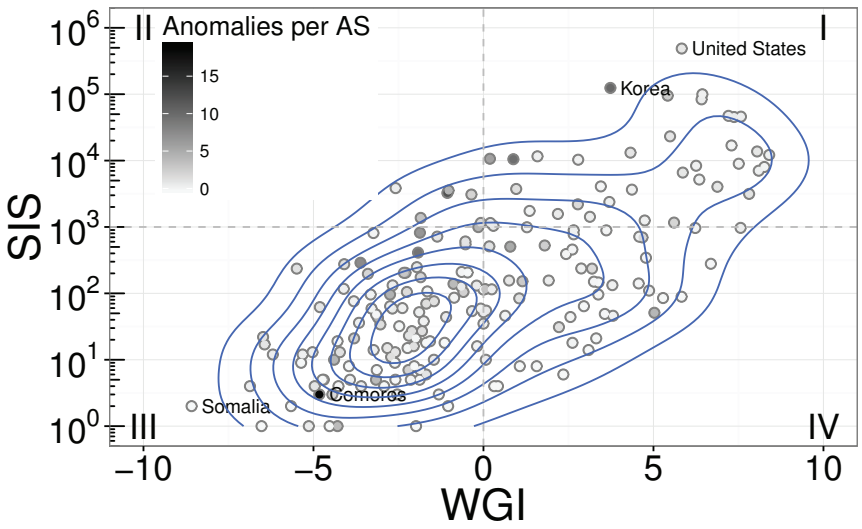

Figure 5.7: Anomaly/ASes ratio in the SIS versus WGI plane for 2011.

5.5 Conclusion

The purpose of this work is to use macroeconomics, unsupervised learning, and theories of crime to find correlations with control-plane anomalies. The motivation of this work is to investigate if solutions to control-plane anomalies should consider incompetence, crime, and national intelligence activities as possible sources. Information about the purpose of attacks can inform the design of defenses. For example, a solution that creates additional complexity for network engineers may increase anomalies caused by failures in human expertise. Alternatively, any solutions which embed national governments as roots of trust may be less than ideal if nation states are the source of these events. Our results reject incompetence as a systematic cause, but we cannot reject crime and national intelligence activities as possible explanations.

We have analyzed the routing anomalies identified by Argus and mapped to the jurisdiction of origin. We have evaluated the statistical significance of macroeconomic indicators corresponding to three theories of ecrime and those that correspond to likely technical expertise for each jurisdiction. This study has as a primary focus an empirical understanding of the variance of anomaly production at the cross-country level. Some limitations need to be taken into account before further discussing the results. The model relied on multiple linear regression through OLS, but OLS

cannot identify the underlying mechanisms for the regression estimates. We also do not know why variables appear more predictive in some countries than in others. The results suggested by the statistical model are based on correlation, which is not an indicator of causality. In that respect, the aggregation and analysis of more features can partially but not wholly improve the understanding of causes in future research. We have offered technical competence, theories of ecrime, and governmental action as candidates for explanations of these correlations. These theories are well-grounded in the literature, motivating their selection, yet the results are in no way deterministic.

We identified some patterns of interest in the data beyond statistical significance that could be partially explained by macroeconomic indicators. First, the overall results indicate that the concentration of anomalies originated in the U.S. is consistently the highest. This suggests that a solution that is deployed in the U.S. may have a disproportionate impact in the reduction of routing anomalies. It also indicates that policies which apply only in the U.S. may be effective despite the global nature of the Internet. The potential for high impact of technology and policy solutions is reified by observation of the two highlighted clusters. Second, we observed significant incidence of anomaly production in countries with poor governance practices and limited investment in security. Among these countries with a striking rate of anomaly production and poor government indicators, Syria is the only country that is highlighted during the entire observation period.

Focusing further on the cluster of countries with poor governmental practices, we identify that indicators of conflict occur more strongly than indicators of poverty, i.e., it is not all about economic deprivation or lack of technical expertise. This is revealed as we track the evolution of the WGI for countries in this cluster. Moreover, as we have shown in the analysis, the four dimensions of governance are not independent of one another [Kaufmann et al., 2011]. In other words, although these variables measure different aspects of governance, they tend to be closely interrelated in a particular country. This is an important consideration because the original formulation of the WGI leverages other dimensions of governance such as voice and accountability, political stability, and absence of violence. This implies that the WGI indicator that we presented in this study is capturing some component of conflict in the highlighted countries. In other words, WGI

can indicate surveillance of a nation's own people.

We found that the distribution of the number of anomalies is well-modeled by either a power law or log-normal distribution. As has been pointed out by [Clauset et al., 2009], in terms of data generation, it is arguably more reasonable to assume that the data are created by a log-normal process because of the distribution of ASes globally and within countries. The type of mechanisms that can be expected to produce these heavy-tailed distributions identifies that power law and log-normal distributions are quite naturally associated [Mitzenmacher, 2004]. In particular, the generative mechanism behind hijacks has been attributed to a multiplicative process in this case, this means the more ASes, the more we expect anomalies. This observation has also been discussed by [Edwards et al., 2015]. Applicably, a multiplicative process is one that can be characterized by Gibrat's rule of proportionate growth stating that the proportional rate of growth of a firm is independent of its absolute size [Samuels, 1965]. Thus, we can apply the analogy to model the size of the countries based on economic, guardianship, and availability factors. It is reasonable to expect the number of hijacks generated in a given country to be proportional to that country's macroeconomic indicators.

Overall, our analysis offers surprises in terms of the role of macroeconomic indicators in anomalies. Much, however, remains to be done. For example, it will be interesting to investigate whether consistent patterns in terms of anomaly production are observable at the level of ASes and ISPs. This would allow us to better understand the microeconomic factors that are responsible for the variance of anomaly production around the globe.

Given that the statistical model relies on assumptions concerning the generation of observed data, the conclusions are also based on the quality of data that we have. It might be the case that the data is incomplete or prone to false positives. Thus, additional analyses on other data sources should be compared with the results here. If this analysis proves predictive over multiple years such that changes in the independent variables are leading indicators in changes of anomalies, then we could have some basis to claim causation.

Our analysis of BGP routing anomalies lends insight into hijacks as a geographical, eco-

nomic, political, and technical challenge. We focused on the analysis of longitudinal cross-country macroeconomic data and its relationship with the frequency of produced routing anomalies. To do so, we collected a data set about BGP incidents from June 2011 to December 2014 and combined this with data from the World Bank. We provided statistical characterizations of the distribution of hijacks over the past four years. The higher–level research question is, whether these are crimes, and if so, can we characterize these empirically as attacks? The specific hypotheses we address focus on determining the economic factors that explain the variance in the number of hijacks originating from different countries.

There appears to be a high and consistently increasing number of BGP incidents, specifically those that are potential hijacks. If we assume that the lack of data for 2014 is just that—a lack of data—and not a lack of incidents, we can expect hundreds of thousands more incidents in the coming years. Although there are false positives, these incidents are not route leaks and cannot be reasonably assumed to be the result of simple technical incompetence. In fact, our analysis found no systematic evidence that these anomalies are a result of incompetence measured by exports of information and communications technologies. (We have no expectation that this would hold for route leaks, which we did not explore.) Other than errors, two remaining explanations are crime and national intelligence. Both of these can be quite difficult to observe. They can also be difficult to distinguish. We found evidence for both possibilities.

To answer the question if BGP attacks are crimes, we use the tools of criminology informed by previous work in empirical analyses of ecrime. Specifically, we built on previous macroeconomic work in online crime using the factors that have been found to be significant in predicting malware, spam, and crowdsourced criminal labor (e.g. CAPTCHA solving). We created a model and evaluated data over time to determine if variables associated with routine activity theory (secure Internet servers, fixed broadband Internet subscribers), economic deprivation theory (GDP per capita by PPP), or structural theory (governance and education) were significant over multiple years. We found that one variable for two of the three theories was significant. Specifically, secure Internet servers and density of fixed broadband Internet subscribers were significant, supporting

routine activity theory. Governance indicators were significant, supporting a structural theory of crime.

We found evidence for the possibility that BGP anomalies (including hijacks) are a new method of ecrime, one that is detected tens of thousands of times each month. If BGP hijacks are a common ecrime vector and follow the patterns of other online crime, the misdirection and filtering of information may be commoditized and become an industry.

A further cluster analysis showed mixed evidence for crime and national intelligence. South Korea and the U.S. are known for being targeted by specific types of ecrime. South Korea is notable for its high level of virtual goods and thus is a target for virtual theft. The U.S. is consistently highly ranked as being the most popular target and the most popular source of common online crime, such as phishing and spam. The U.S. maintains its role as an outlier in this analysis; however, South Korea has a greater density of anomalies per AS. Although South Korea is itself characterized by high levels of governmental transparency and competence, it remains technically at war with North Korea, as there is no peace agreement. North Korea has been indicated in virtual as well as physical attacks against South Korean nationals. Similarly, although the U.S. is transparent, wealthy, and competent by global measures, the U.S. has not only never recognized North Korea as a nation, it also is actively involved in warfare in Afghanistan as well as having a wide range of additional military engagements. Given the extremely large investment in national intelligence and national defense as well as the current involvement in conflicts overseas by the U.S., plus the situation with respect to North Korea, their presence in anomalies can also be seen as supporting the national intelligence explanation.

For the other cluster, that of nations with low numbers of secure Internet servers, the increase in density of anomalies per operating autonomous systems shows a striking overlap with nations with civil disputes over the time period studied. The relatively small number of countries prevents meaningful statistical analysis of these few nations, but the overlap between national conflict and anomalies is striking.

In summary, our findings provide statistical evidence for one of the three proposed explanations

for anomalies: crime. We found weaker evidence in terms of governance and clustering for a second explanation: national intelligence. We found no evidence for incompetence per se.

6 Characterizing Routing Anomalies Through Graph Mining

"We must be careful not to believe things simply because we want them to be true. No one can fool you as easily as you can fool yourself."

— Richard P. Feynman

6.1 Introduction

In this chapter, we address the challenge of early identification of routing anomalies using an anomaly detection approach. In particular, we conduct a case-based systematic analysis of the changes in the topology of AS-level graphs that are associated with three very well-known large-scale incidents. The three cases we address were easily identified following the large-scale disruption but not before. We show that before the incidents there are changes in the topology of the graphs but there are not statistically significant. Here the results suggest that further research is needed to support the idea of using AS-level graphs for early identification of routing anomalies.

6.2 Problem

The Internet—one of the most beautiful man-made complex systems—has shown to be under constant evolution [Zhang et al., 2008, Edwards et al., 2012]. The Internet is composed of ASes which are autonomous entities with their own routing policy and administrated by a single entity. ASes exchange traffic between them by means of BGP [Rekhter et al., 2006]. The interchange of routes between ASes allows them to know the global routing status of the Internet at any time. This dynamic system can be represented by temporal graphs in which the nodes are the ASes and the BGP peering relationships between them are the links. There is extensive literature studying the structure of the Internet at the AS-level for specific time snapshots [Faloutsos et al., 1999, Albert

et al., 2000, Zhou and Mondragón, 2004]. More recent work also focuses on the dynamics of the topology of the Internet [Zhang et al., 2008, Edwards et al., 2012]. However, to our knowledge, there are no quantitative studies in the literature that systematically evaluated the effect that large-scale routing anomalous events have in the topology of the Internet at the AS-level.

The purpose of this chapter is to quantitatively study the topology of the Internet when real-world anomalies have been injected into the system. Anomaly detection techniques rely on conducting measures on the control-plane level (using BGP feeds) or data-plane level (by exploring reachability of IP addresses in suspicious announced routes), or a combination of both. Anomaly detection schemes do not require changes in the protocol itself. They primarily are used in detecting anomalies based on passive or active measurements, i.e., to alert operators to mitigate threats [Shi et al., 2012, Khare et al., 2012, Zhang et al., 2010]. However, in many of the anomaly detection schemes, the prefix measurements are precomputed and not dynamic. This implies that their underlying mechanisms need to be recomputed if there is a change in the observed routing infrastructure.

Anomaly detection techniques have been shown to be able to identify anomalies under certain conditions. However, it is still an open issue to understand and characterize the occurrence of anomalous events on the Internet using the information extracted from the reachability graphs built from BGP announcements. Here we seek to contribute to the understanding and characterization of routing incidents through the use of methods from other fields. Specifically, we use graph mining from the study of human social networks to detect these anomalies. Not only is a graph-theoretical approach suitable for the connection of ASes, this approach can incorporate the dynamic behavior of the observed routing infrastructure.

For that purpose, we conduct a case-based systematic analysis of the changes at the topological level of the AS graphs that are associated with three very well-known anomalous events. The three cases we address were easily identified following the large-scale disruption but not before.

Our methodology is grounded in building the AS-level graphs and analyzing the metadata that is involved with well known incidents. Our work is differentiated from the work in [Kruegel

et al., 2003, Gaertler and Patrignani, 2004] in that we use dynamic update information from the RouteViews project (with granularity every 15 minutes) to reconstruct the network topology at the AS-level and study the robustness of network topological properties, before, during, and after the incident. This approach aims for differentiation between normal behavior at the network level and disruption or anomalous changes during the incidents. From this analysis we expect that topological signatures from the AS-level graph representation can be used to infer when an anomalous routing event is happening before widespread disruption.

6.3 Methods

Every AS originates the prefixes that have been allocated to it. All ASes can announce prefixes as well as paths. ASes build a graph of interactions with others ASes based on information about reachable paths to IP prefixes. The reachability of these paths is determined through BGP announcements that ASes receive from their neighbors. Gateway routers in the ASes use route updates to modify their routing tables, and these determine how to direct traffic. It has been shown that routing decisions depend mainly on path length (i.e., the number of hops to reach the destination prefix); secondarily on the cost of directing traffic through a specific neighbor based on previously established business contracts; and then on diverse tertiary criteria [Goldberg, 2014]. The resulting graph is a dynamic system based on protocol incentives and economic constraints with continuous addition and deletion of nodes and edges [Pastor-Satorras and Vespignani, 2007].

Using the graph of ASes, we performed a longitudinal empirical analysis of the network topological measures that correlate with the occurrence of three major BGP disruptions. For each event, we reconstructed the evolving network topology around the date and time of these incidents. This procedure is done to identify statistically significant changes in the graph topology.

6.3.1 Data Sources

In this section, we detail the data sources (along with their preprocessing) that were used to perform the analysis which follows. We start by describing the data used for the construction of (i) the

database of the large-scale routing anomalous events (to establish the ground truth); (ii) the AS-level data; (iii) the description of the AS-level graph representation; and then end with (iv) the definition of the graph topological properties measured in the Internet graphs.

6.3.1.1 Routing Anomalous Events

We considered the AS-level graphs for very well-known cases of routing anomalous events that led to large-scale disruptions during the last few years. In particular, we performed an analysis of (i) an Indonesian ISP hijack that covered much of the world; (ii) an Malaysian ISP that generated global collateral damage by leaking prefixes to large-scale providers; and (iii) an Indian ISP that hijacked prefixes of other important Internet players. Note that these anomalous events have been studied and corroborated from different sources. We describe more details about these incidents below. Events are listed in chronological order.

An Indonesian ISP hijacks the world

On April 2, 2014, starting at 18:26 UTC, Indosat (one of the largest telecommunications providers in Indonesia) announced more than 320 000 IP prefixes belonging to other networks. In fact, Indosat announced roughly two-thirds of the entire Internet address space [Zmijewski, 2014]. A large fraction of the hijacked prefixes belonged to Akamai, which is one of the larger Content Delivery Networks. This event lasted for several hours until approximately 21:15 UTC. Traffic continued to be delivered; however, the path of the traffic was significantly perturbed.

Global collateral damage of the Telecom Malaysia leak

On June 12, 2015, starting at 08:43 UTC, Telecom Malaysia announced about 179 000 IP prefixes to Level 3 (the largest crossing AS) [Toonk, 2015a]. Level 3 accepted these announcements and then propagated the routes to their peers and customers around the word. Because Telecom Malaysia is a customer of Level 3, the routes announced by Telecom Malaysia were identified as a preferred delivery route for Level 3. At around 10:40 UTC, there were slowly observed improve-

ments, and by 11:15 UTC the errors in the RIB began to be resolved. Note this was a leak, so the data were not delivered after being transmitted to Telecom Malaysia.

Large-scale BGP hijack in India

On November 6, 2015, starting at 05:52 UTC, Bharti Airtel Ltd., claimed the ownership of about 16 123 IP prefixes. These prefixes corresponded to more than 2000 unique ASes [Toonk, 2015b]. This event became widespread because two large ASes (e.g., Cogent Communications and GlobeNet Cabos Submarinos S.A.) accepted and propagated these routes to their peers and customers. Legitimate owners of the prefixes included Akamai, Tata Communications, and Apple Inc. This event lasted until approximately 14:40 UTC.

We summarize the details of the incidents used in this study in Table 6.1.

Table 6.1: Summary of large-scale routing incidents.

Incident	Date	Start time	End time	Duration
Indosat	2014-04-02	18:26 UTC	21:15 UTC	$\approx$ 2.9 h
Telecom Malaysia	2015-06-12	08:43 UTC	11:15 UTC	$\approx$ 2.7 h
Bharti Airtel Ltd.	2015-11-06	05:52 UTC	14:40 UTC	$\approx$ 8.9 h

6.3.1.2 BGP Data

We collected BGP measurement data using BGPStream[1]. BGPStream provides an open-source software framework for the analysis of historical and real-time BGP data [Orsini et al., 2016]. To do so, BGPStream extracts data directly from route collectors. A route collector is a host running a collector process. The collector emulates a router that establishes BGP peering seasons with real BGP routers. These collection points are known as vantage points (VPs, hereafter).

A BGP router maintains its reachability information in the Routing Information Base (RIB). Together, the VPs ideally provide a list of paths between Autonomous System Numbers (ASNs) of every reachable network. The collection points aggregate data including update messages that

[1]Available at https://bgpstream.caida.org/

reflect routes being added or deleted from the RIB. Update messages contain fine granular information about routing dynamics [Mazloum et al., 2014], i.e., changes in the paths. By sampling changes in the routing table of the VPs, collectors can reconstruct RIBs from their peering routers. This constitutes a partial view of the Internet at the AS-level, i.e., an undirected graph in which vertices are ASes and the edges are routing links between them.

There are two popular projects running route collectors processes, RouteViews [Meyer, 2004] and RIPE RIS [RIPE NCC, 2011]. They make dumps available in public archives. At the time of this writing, they operate 19 and 17 collectors, respectively, which peer with hundreds of VPs respectively [Gregori et al., 2012]. RouteViews and RIPE RIS collect a RIB dump every two hours and eight hours, and update dumps every 15 and five minutes, respectively.

BGPStream allows the setup of different parameters in the data collection process. In particular, it is possible to manipulate the start and end date of data collection, and the specific project running route collector processes, among other variables. We collected approximately two days of observations around the start date of each of the incidents in Table 6.1—to gather observations before, during, and after the selected events. The purpose of collecting data over this time period is to be able to distinguish between regular and anomalous behavior. In this analysis, we only collected BGP measurements from the RouteViews project. Previous research has shown that there is a considerable overlap between the measurements from RouteViews and RIPE RIS projects [Chen et al., 2009]. Thus, it is not expected that by adding observations from RIPE RIS, we are augmenting the resolution of the observed graphs.

6.3.1.3 AS-Level Graph Representation

We aggregated every possible path in the RIBs among collectors (at a certain time) to build snapshots of the Internet topology. The resulting graph topology was constructed from observed paths derived from BGP updates. For repeated paths among the different collectors, we only considered one instance of the path. Specifically we modeled the AS-level topologies as graphs with the following considerations.

Consider the sequence of n intervals $A = \{A_1, A_2, \ldots, A_n\} = \{A_t\}_{t=1}^{n}$, where

1. $A_t = [a_t, a'_t)$ for all $t < n$ and $A_n = [a_n, a'_n]$ for $t = n$; and

2. $a_t < a'_t = a_{t+1}$ for all t;

An interval represents a fixed-length unit of time, i.e., the granularity at which BGP updates dumps are collected. In this case, it is 15 minutes, based on the data available from the RouteViews project. Note that dump times are synchronized among the collectors for each experiment discussed in this paper. Condition (1) implies that all intervals are left-closed and right-open (except the last one which includes a'_n). That is, if events at time t are part of one sequence, then only events later than t are part of the next sequence. It guarantees that the sequence of intervals is disjoint. Condition (2) implies that intervals are non-empty. Note that a'_t and a_{t+1} represent the time instants of a transition between intervals. For any interval A_t, the right endpoint a'_t corresponds to the left endpoint of the interval A_{t+1}. Together with Condition (1), Condition (2) guarantees that the union of all intervals $\bigcup_{t=1}^{n} A_t = [a_1, a'_n]$ is a closed interval.

In addition, we let $\mathcal{H} = \{1, 2, \ldots, N\}$ be the set of nodes (e.g., set of ASes). Then, $\mathcal{V}(t) \subseteq \mathcal{H}$ is the subset of nodes which interact (i.e., which have an identified path during interval $A_t = [a_t, a'_t)$). Let $\mathcal{E}(t) = \{e_{ij}(t) : i, j \in \mathcal{H}\}$ be an adjacency matrix of edges $e_{ij}(t)$ that captures the existence of a routing link between node i and node j during interval A_t. Let the graph $\mathcal{G}(t) = (\mathcal{V}(t), \mathcal{E}(t))$ represents an undirected graph that captures all interactions that occur from endpoints a_t to a'_t, $t \in \{1, 2, \ldots, n\}$. The total number of nodes of nodes in $\mathcal{G}(t)$ is $N(t) = |\mathcal{V}(t)|$. The sequence $\{\mathcal{G}(t)\}_{t=1}^{n}$ denotes the graph series G.

Let $P_{ij}(t)$ be a set of paths between nodes i and j at time interval t. Each path $p_{ij}(t) \in P_{ij}(t)$ is a sequence of edges made from nodes from $\mathcal{V}(t)$ that forms an undirected path that starts at i and finishes at j. For example, $\{(i_1, i_2), (i_2, i_3), \ldots, (i_{k-1}, i_k)\}$ is a path where $i_1 = i$ and $i_k = j$, and each node in the sequence $i_1, \ldots, i_k$ is distinct. Let $p^*_{ij}(t)$ be the shortest (geodesic) path between nodes i and j at time interval t. Let $d^*_{ij}(t)$ be the length of the shortest path between nodes i and j, i.e., $|p^*_{ij}(t)|$.

Finally, let $q_i(t)$ represent the set of neighbors of node i at time t, i.e., $q_i(t) = \{j : e_{ij}(t) \in \mathcal{E}(t)\}$. Then, the degree of node i is $|q_i(t)|$ and the total number of edges of $\mathcal{G}(t)$ is $E(t) = \sum_{i \in \mathcal{V}(t)} |q_i(t)|$.

Using these conditions and this approach, we constructed graphs as each time period. That allowed us to measure the topological properties of the graph as the anomalies were introduced to the network, diffused, and then removed.

6.3.1.4 Topological Properties

We measured a set of graph topological properties to study their correlation with very well-known cases of large-scale Internet disruptions. The selected topological measures are the most relevant for the understanding of the structure and function of the Internet according to [Edwards et al., 2012, Haddadi et al., 2008]. We grouped these properties in three different categories: global structure, path length, and community structure measures.

Table 6.2 summarizes each topological category and their relationships with the structure and function of the Internet. We discuss the details about these categories below.

Table 6.2: Summary of graph topological properties and its relationship with Internet's performance.

Property	**Internet effect**
Global structure	ASes connectivity (importance)
Average path length	Routing efficiency
Clustering	Peering structure (alternate routes)

Global structure

The global structure of a graph has been used to characterize the state of a graph in a certain snapshot. In the context of the Internet, it has been shown that these measures are relevant to understand ISP regulation [Hofmeyr et al., 2013] and the robustness of the Internet [Doyle et al., 2005].

Number of nodes: This measure corresponds to the number of unique nodes in the graph at a certain time. In this chapter, we have reported the number of nodes as $N(t)$.

Number of edges: This measure corresponds to the number of unique edges in the graph at a certain time. In this chapter, we have reported the number of edges as $E(t)$.

Maximum degree: The degree of a node is the number of edges attached to it. In this chapter, we report the maximum degree across the nodes in the graph as $\max\{|q_i(t)| \; \forall i \in \mathcal{V}(t)\}$.

Path length

The average path length relates to the number of hops between nodes for every possible pair. In the context of the Internet, it is an important property to study because it relates to the efficient routing of packets. Although it is known that not all packets travel through the shortest path given commercial agreements, routing efficiency is ultimately influenced by shortest path measures. Here, we used the average path length, i.e., the mean of the shortest paths between each pair of nodes in the graph. The shortest path between two nodes belonging to different components is said to be infinite. More formally, it is defined as $L(t) = \frac{1}{N(t)(N(t)-1)} \sum_{i,j \in \mathcal{V}(t)} d^*_{ij}(t)$.

Community structure

Community measures relate to the likelihood of finding a group of nodes that are able to form substructures. In the context of the Internet, it has been shown that community structure is key to understand the tiered structure of the Internet [Yang and Leskovec, 2014] and its resilience to both random or targeted removal of nodes [Doyle et al., 2005]. We defined each of the community structure measures used below.

Average clustering coefficient: The local clustering coefficient of a node $i \in \mathcal{V}(t)$ quantifies how close its neighbors are to be connected as well. In particular, the local clustering of node i is defined as $\gamma_i(t) = \frac{2|e_{jk}:\, j,k \in q_i(t),\, e_{jk}=1|}{|q_i(t)|(|q_i(t)|-1)}$. In this paper, we reported on the average clustering coefficient defined as $\bar{C}(t) = \frac{1}{N(t)} \sum_{i \in \mathcal{V}(t)} \gamma_i(t)$.

Components: A component is a subgraph in which any two pair of nodes are connected to each other by at least a path. In this paper, we reported on the average size of the components.

Within each category, we have also made use of group techniques on the vertices to decompose the graphs into shells—informed by previous empirical observations about the actual structure of the Internet [Alvarez-Hamelin et al., 2008]. In particular, we tangentially applied k-shell decomposition to analyze properties from each of the previous described categories.

k-shell decomposition

The k-shell decomposition fragments a graph between core and crust subgraphs respectively. The k-shell of a node (also called coreness) is a measure of the centrality of the node with respect to its neighbors. This decomposition technique is an iterative process. It starts from degree $k = 1$, and in every step, nodes with similar degree are removed until the nucleus of the graph is revealed—the maximal k that keeps $N(t)$ larger than zero. The details about this process at time instant t are described below.

Step 1: Compute the adjacency matrix $\mathcal{E}(t)$ and identify nodes with degree $k = 1$, i.e., $\{i \in \mathcal{V}(t) : |q_i(t)| = 1\}$.

Step 2: Remove all nodes with degree equals to k, i.e., $\{i \in \mathcal{V}(k) : |q_i(k)| = k\}$. This results in a pruned adjacency matrix $\mathcal{E}'(t)$.

Step 3: Compute the degree of each node from the remaining set of nodes. If there are nodes with degree equals to k, step 2 is repeated—producing a new adjacency matrix $\mathcal{E}'(t)$. Otherwise, return to step 1 with an increased value of $k = k + 1$ and $\mathcal{E}(t) = \mathcal{E}'(t)$.

At any given time t, the k-shell is made of all removed nodes (and their respective edges) in a given degree (step) k. The k-shell decomposition reveals hierarchies of ASes. The subgraph that is generated from the accumulation of all removed nodes, i.e., in all previous $k-1$ shells, is called the k-crust [Lahav et al., 2016]. In the context of the Internet, the k-crust reveals the periphery of the AS-level graph. The subgraph that is formed from the remaining graph at any given step t is called the k-core—the maximum subgraph with minimum degree at least k. In general, the nucleus (the k-core graph with the largest possible k) of the Internet has been studied with the aim of understanding the evolution of the Internet [Zhang et al., 2008]. It is worth noting that at the

end of every k step, new k-shell, k-crust, and k-core subgraphs are produced. More details on this decomposition method can be found in [Pittel et al., 1996].

Figure 6.1 shows the representation of an Internet map at the AS-level—as of the end of the second week of May 2017. The visualization algorithm uses weekly updated data from the Route-Views project. A node's annotations correspond to the country at which ASes are registered. The AS that is tagged in Figure 6.1 corresponds to Indiana University (ASN 87)—based on the origin IP address at which this figure was generated. In addition, the left scale represents the degree of the nodes (node sizes are scaled with respect to their degree). The right scale represents the k-shell index of the nodes (hotter colors represent ASes that are more towards the core). This map is generated according to the algorithm described in [Alvarez-Hamelin et al., 2005]. I include the

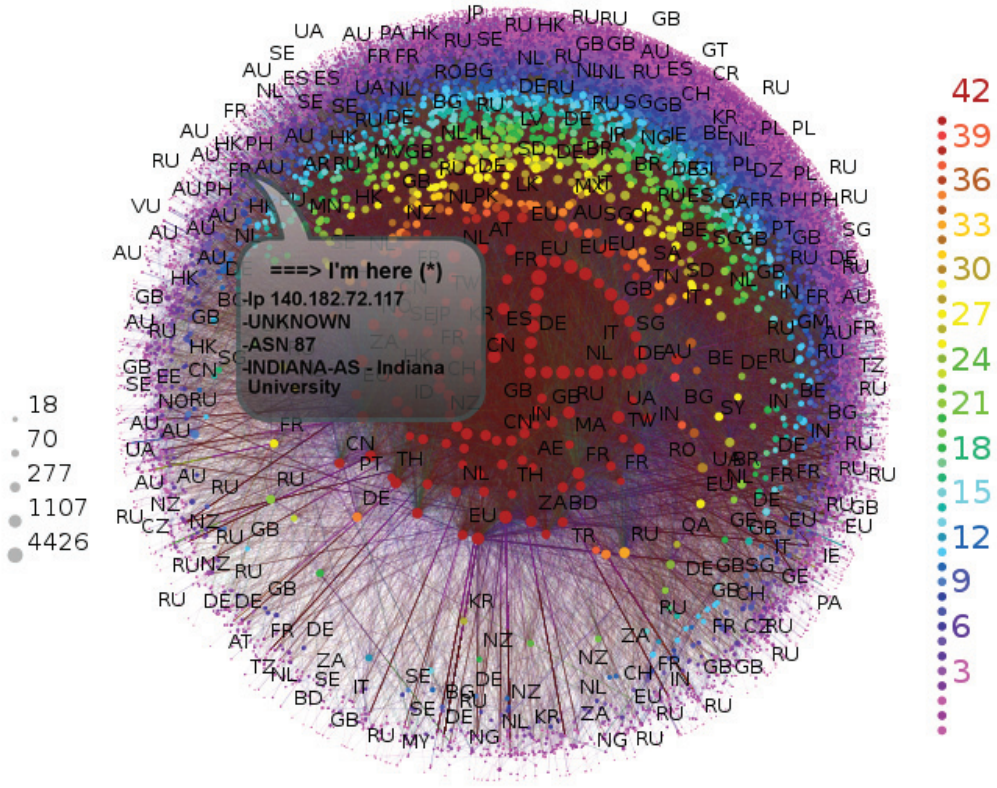

Figure 6.1: Graphical representation of the AS-level graph.

schematic illustrating the hypothesis that we test in Figure 6.2 again from Section 1.

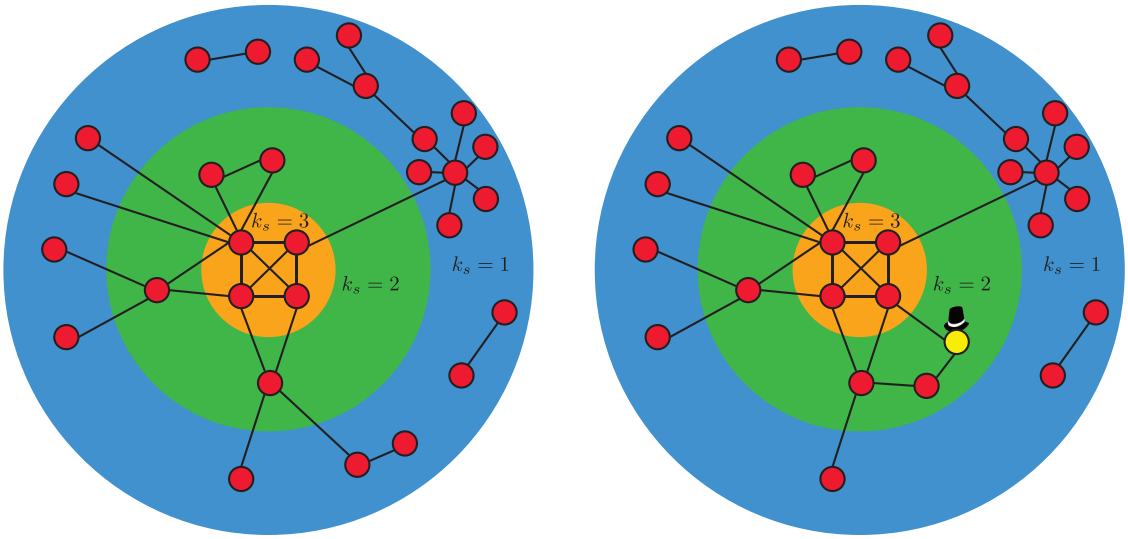

Figure 6.2: A schematic representation of the graph under k-shell decomposition for different values of shell k_s. The AS-level graph under normal conditions is on the left and the same topology when a BGP hijack is committed is on the right. The hijacker is highlighted with the black hat.

6.4 Results

Our results indicate that the anomalous routing events are correlated to some extent with changes in some properties in the graph topology generated from the BGP measurements. The main idea is to examine the structure of the network before, during, and after the three large-scale anomalies. Topological changes in the network at the AS-level will allow us to test the hypothesis as to whether dynamic transitions in the network structure can be used as a detection signature of anomalous Internet routing events.

We report the results based on the analysis of three large-scale routing incidents using BGP updates from BGPStream. In this section, we have summarized the properties of each reported time series based on the category to which these belong and the corresponding anomalous routing incident. Remember that we focus on properties that are related to the structure and the function of the Internet. These properties produce a single deterministic number when evaluated and are not distributions.

We used the same visualization conventions for every plot. First, we used open circles when plotting the raw data. Recall the raw data corresponds to the empirical measures for each graph

topological property. Second, we used moving average to smooth out short-term fluctuations and highlight long-term trends in the time series. Solid lines represent the moving average when using a centered window of 10 observations around each data point. We use a small number for the length of the rolling window to avoid smoothing the signal aggressively and not being able to pinpoint the anomalies. Third, we computed the moving standard deviation using similar parameters as used for the moving average. We highlighted the region comprehended between $\pm$ one standard deviation for each data point in faint red, i.e., the band of statistically significance. Finally, we highlighted in yellow the beginning and end times of each of the incidents described in Table 6.1. We marked minor ticks in the time axis to represent intervals of 15 minutes, i.e., the time between consecutive BGP updates from the RouteViews project.

6.4.1 Global Structure Measures

An Indonesian ISP hijacking the world

Here, we report on the results of centrality measures which illustrate the prominence of ASes. Figure 6.3 shows the maximum degree of each graph snapshot during the observation period. From this plot, it is possible to infer that the only significant changes in this measure are for the graphs captured at April 2, 2014, at 12:00 and April 3, 2014, at 6:00. These discontinuities are consistent with the ones illustrated in Figures A.1 and A.2 (see Appendix A.1).

Figure 6.4 shows the number of nodes at various k-levels of crust graphs, i.e., $k = 1, 10$, and the maximum k possible—the one that encloses the biggest crust of the Internet. As we might expect, it is not possible to observe significant changes regarding the total number of nodes in this time series. This suggest that the periphery graph keeps almost the same with respect to the number of ASes.

Global collateral damage of Telecom Malaysia leak

For this incident, Figure 6.5 shows the maximum degree. As can be seen, the only significant variations for this property occurs for observations derived at June 11, 2015, at 10:00, and 18:00;

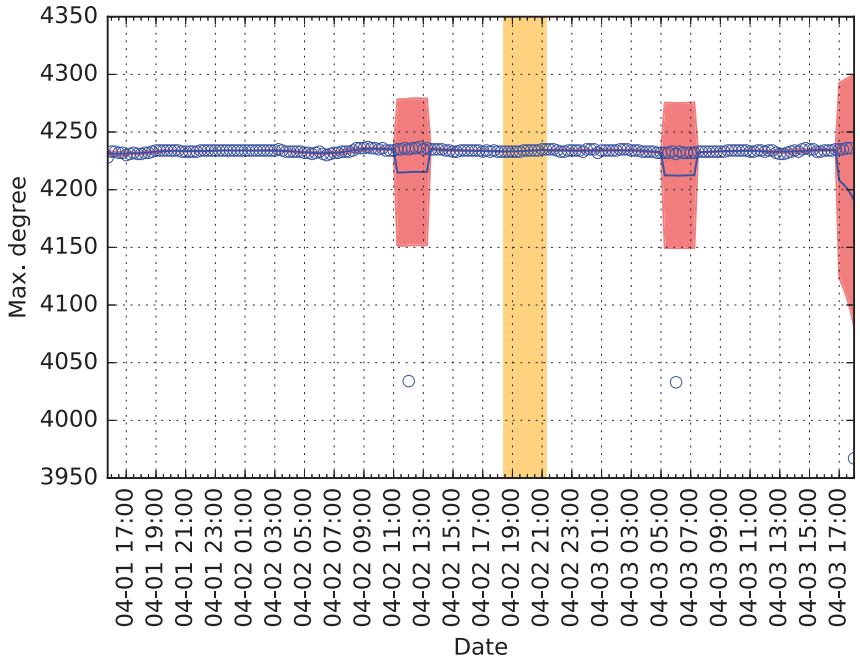

Figure 6.3: Maximum degree Indonesia event.

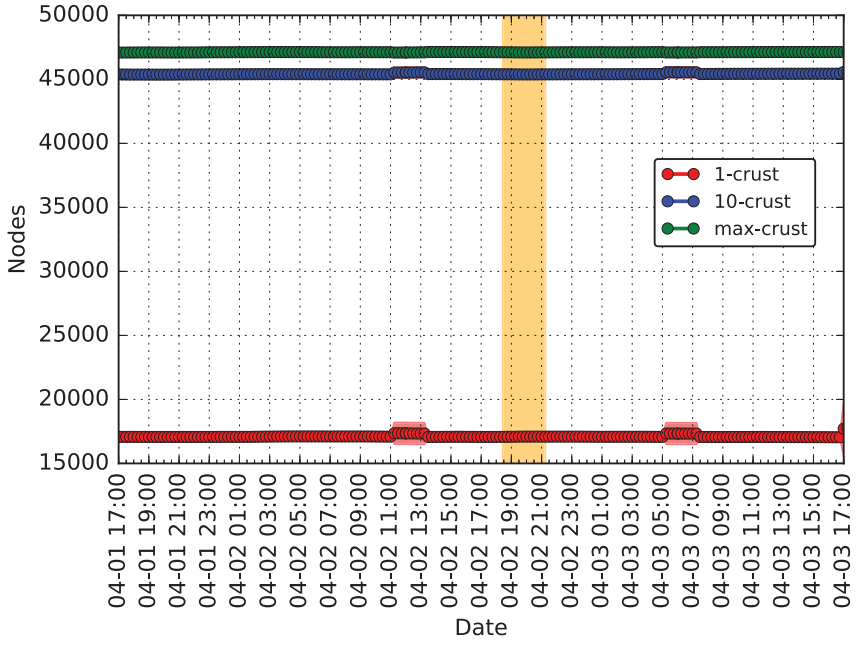

Figure 6.4: Nodes per crust Indonesia event.

June 15, 2014, at 00:00, and 18:00; and June 13, 2015, at 00:00, and 8:00. Similarly, for the crust graphs, we computed the number of nodes as is shown in Figure 6.6. We did not observe significant variations in the number of nodes that is in the largest crust.

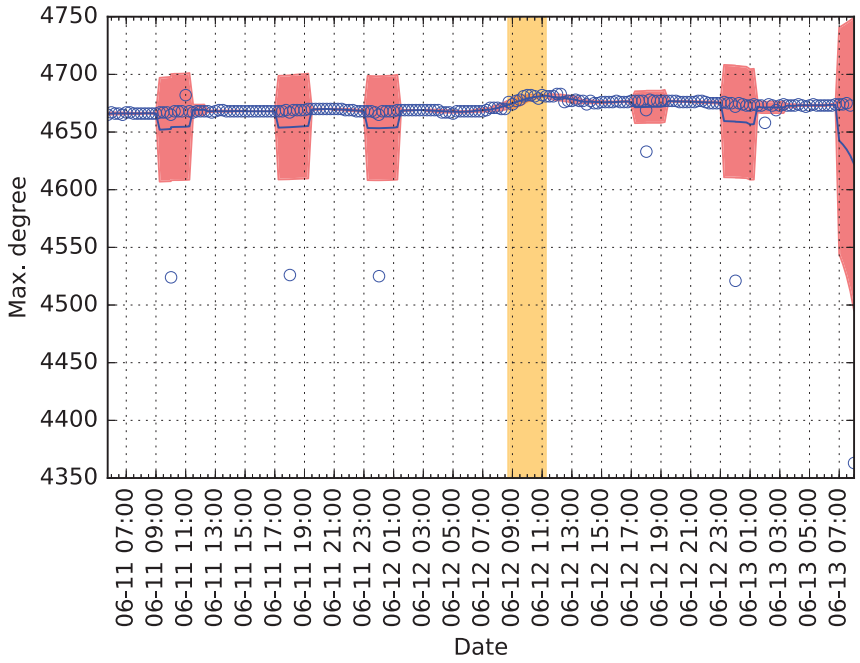

Figure 6.5: Maximum degree Malaysia event.

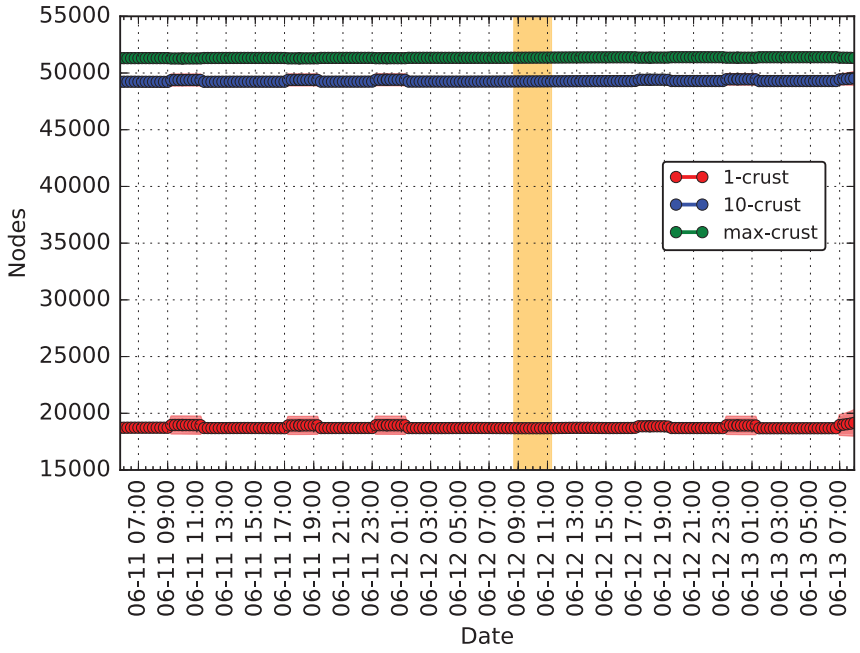

Figure 6.6: Nodes per crust Malaysia event.

Large scale BGP hijack in India

For the Indian incident, we tracked of the same properties we did for the Indonesian and Malaysian incidents. In particular, Figure 6.7 shows the maximum degree for each graph snapshot during the

observation period. As in the previous cases, there are some discontinuities in the time series. The discontinuities are evident in November 5, 2015, at 10:00, 12:00, 14:00, 16:00, 18:00, 20:00, 22:00; and November 6, 2015, at 00:00, 02:00, 04:00, and 06:00. In particular, the discontinuity on November 6, 2015 at 4:00 is more stronger. This suggests that the node with most connections in the graph suddenly decreases its degree—which is remarkable given that it happens almost two hours before the anomaly was reported by other BGP monitoring projects, e.g., BGPmon.net and Dyn Research. Similarly, Figure 6.8 shows the number of nodes in the crust subgraphs. This measure does not reveal significant changes during the observation period.

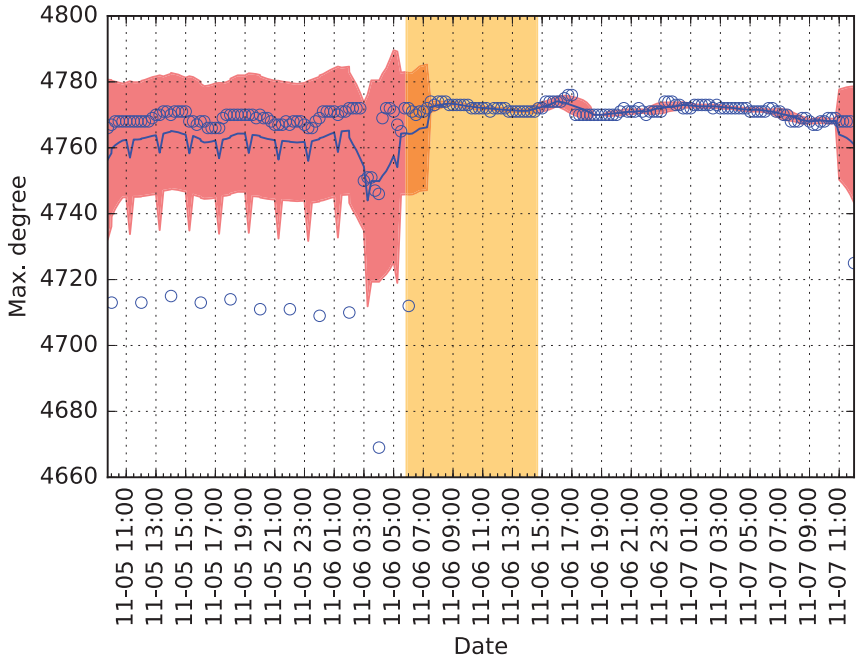

Figure 6.7: Maximum degree India event.

6.4.2 Path Length

An Indonesian ISP hijacking the world

Figure 6.9 shows the average path length for each graph snapshot during the Indonesia incident, i.e., for different values of k in the crust. In particular, we observed that the average path length seems to be altered based on the value of k being analyzed. However, there are no observations

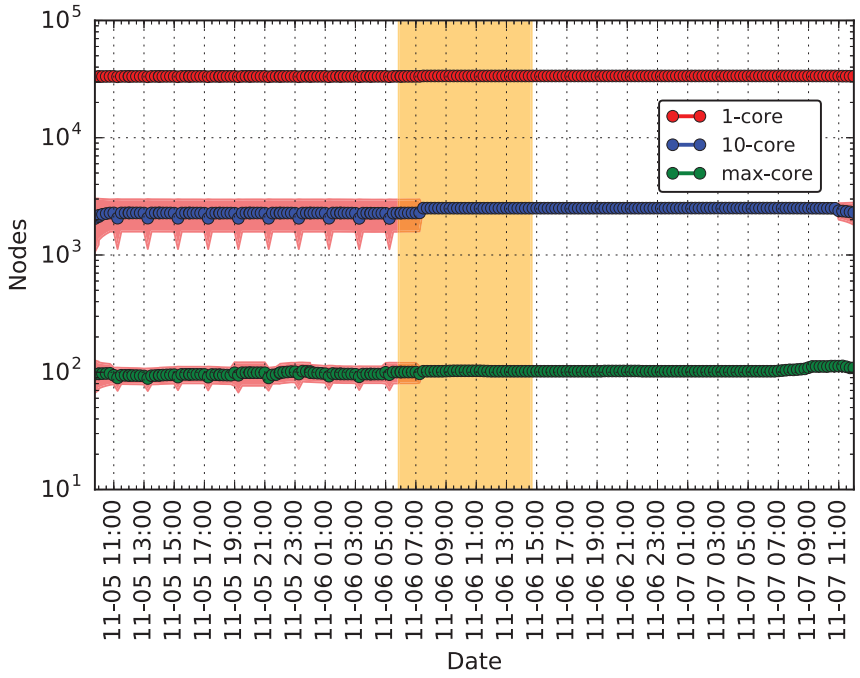

Figure 6.8: Nodes per crust India event.

outside the band of statistically significance. For a value of $k = 10$ this measure tends to decrease in agreement with the discontinuities noticed in Figure A.1, for a bigger crust graph—the whole graph without the nucleus of the network—this measure tends to increase. This is an important observation given that these properties are directly related with the routing efficiency and the number of alternative routes to reach different networks. This might be expected in the case of a disruption of service observed in these types of events.

Global collateral damage of Telecom Malaysia leak

Figure 6.10 shows the average path length measure over different crust subgraphs for the Malaysia incident. As for the Indonesian incident, relative changes in this measure over the crust depends on the k value being analyzed, i.e., the average path length decreased for a value of $k = 10$, but it increased for the the largest generated subgraph. Note however that these variations are not outside the band of statistically significance.

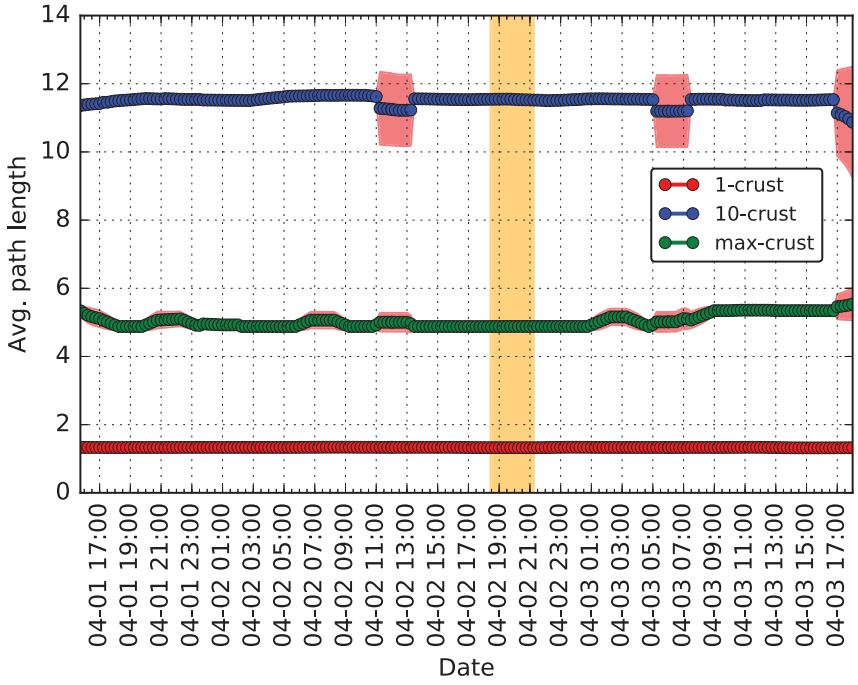

Figure 6.9: Average path length in the crust Indonesia event.

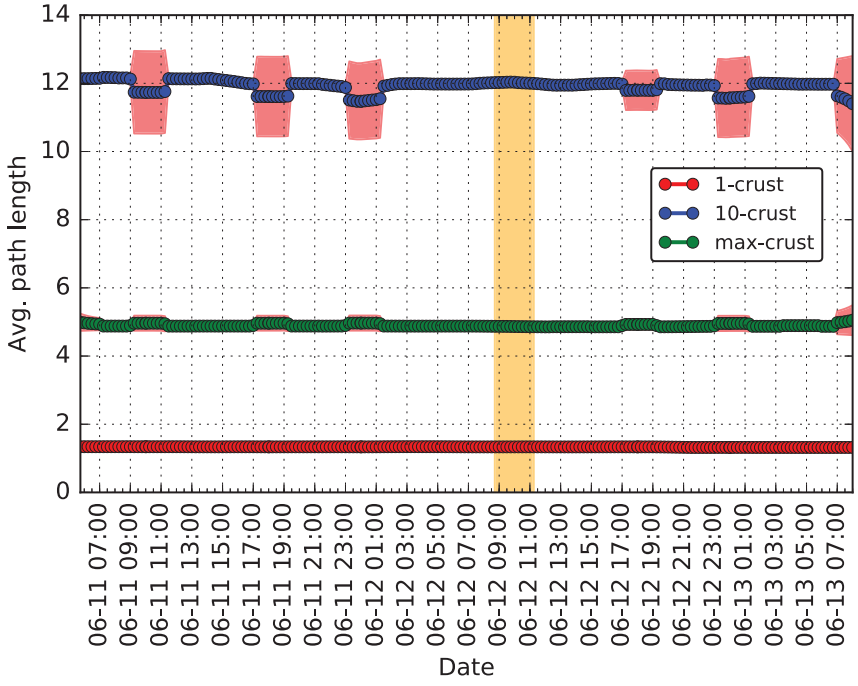

Figure 6.10: Average path length in the crust Malaysia event.

Large scale BGP hijack in India

Similar to Figures 6.9 and 6.10, Figure 6.11 shows a changing pattern in the average path length depending on the crust being analyzed. These changes correlated with the abrupt discontinuities

illustrated in Figure 6.7, but not outside the band of statistically significance.

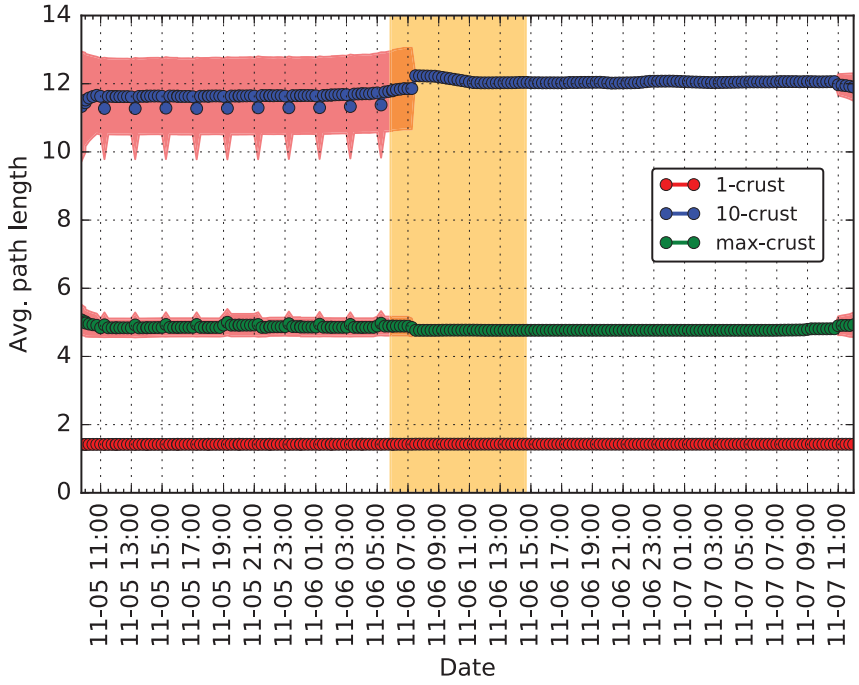

Figure 6.11: Average path length in the crust India event.

6.4.3 Community Structure

An Indonesian ISP hijacking the world

Community structure measures provide a sense of how clustered are nodes in a graph. This is relevant when studying how disruption in the grouping between ASes can be an indicator of an anomalous event. Figure A.10 shows the average clustering coefficient of the graph snapshots during the period of study. This measure seems to be stable during the observation period except for the discontinuities around the same time as we observed before in the centrality and average path length measurements.

We then looked at the average clustering coefficient in the core and crust subgraphs for different values of k in Figures A.11 and 6.12. It is worth nothing that—in advance—of the reported times of the incidents, it is possible to observe some disruptions in the clustering measure for both core and crust subgraphs. More interestingly, Figure 6.13 shows the average size of the components for

the crust subgraphs. There are abrupt changes in the mean size of these components during the observation period, suggesting reallocation of ASes across the multiple shells. Note however that these observations are not outside the band of statistically significance.

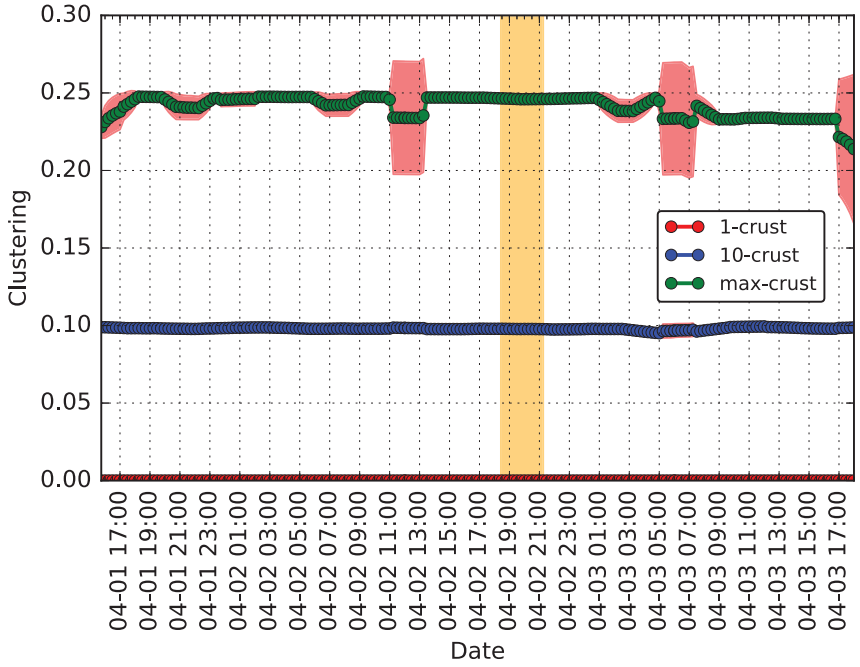

Figure 6.12: Clustering per crust Indonesia event.

Global collateral damage of Telecom Malaysia leak

For the Malaysian incident, Figure A.12, shows the average clustering coefficient for the whole graph—with no k-shell decomposition applied yet. Discontinuities in the signal are observed in correspondence with the same behavior exhibited for other structural properties measured at the general graph, e.g., Figure A.4. We also studied the patterns in the number of nodes in the core and crust subgraphs. Figures A.13 and 6.14 shows the variability in these patterns. They seem to coincide with previous illustrated discontinuities for the whole graph snapshots. Figure 6.15 shows the mean number of nodes in the graph components of the crust subgraphs. Variations in this property are not as much evident as for the case of the Indonesian incident.

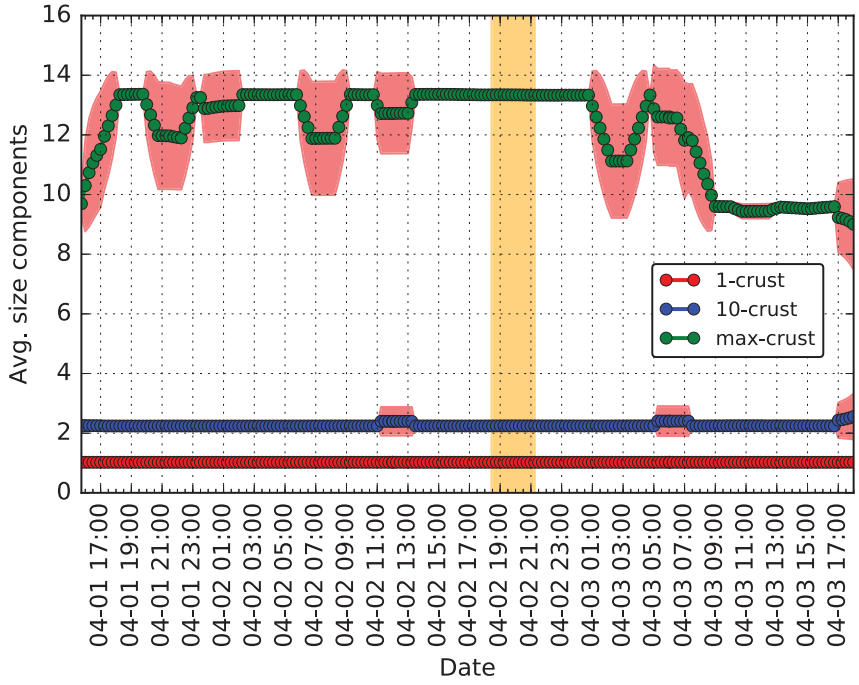

Figure 6.13: Average size components crust Indonesia event.

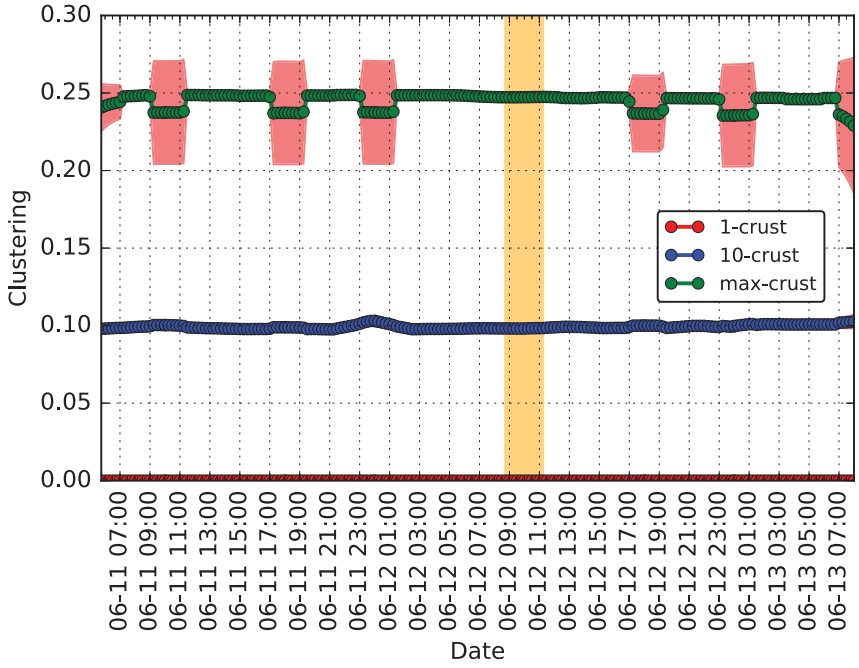

Figure 6.14: Clustering per crust Malaysia event.

Large scale BGP hijack in India

Finally, for the Indian incident, we report similar metrics in the clustering measurements as for previous anomalous events. Figure A.14 shows the time series of the average clustering coefficient.

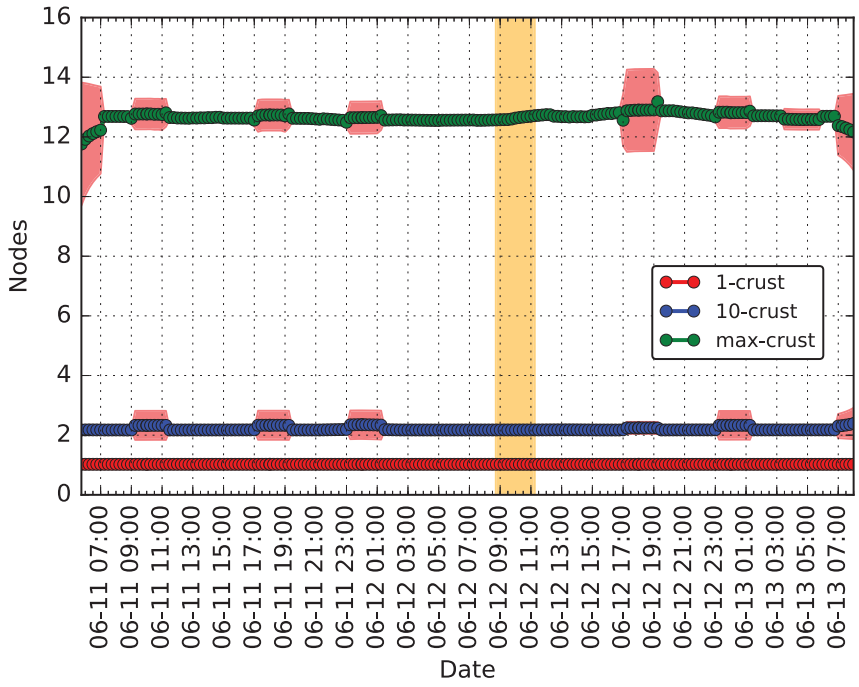

Figure 6.15: Average size components crust Malaysia event.

In general, the signal has discontinuities in accordance with centrality measures plots. Figures A.15 and 6.16 capture the same property for core and crust subgraphs. It is of interest that for both—core and crust—measurements, there is a significant reduction in the clustering even under the presence of discontinuities as noticed in the case of centrality measures. Finally, the time series in Figure 6.17 confirms this observation—when it is observed continuous disruption in the average size of components during the observation period.

6.5 Conclusion

The purpose of this chapter was to explore the applicability of a different approach, one that uses graph mining to the challenge of identifying BGP anomalies. In doing this we consider inputs from the dynamic representation of the AS-level graph. We found some value in examining the robustness of AS-level graph properties in terms of early warning of incidents. A solution that focuses on the monitoring of the dynamic evolution of the AS-level graph may help detect anomalies that are not yet evident using traditional control- and data-plane measurements. In the k-shell decom-

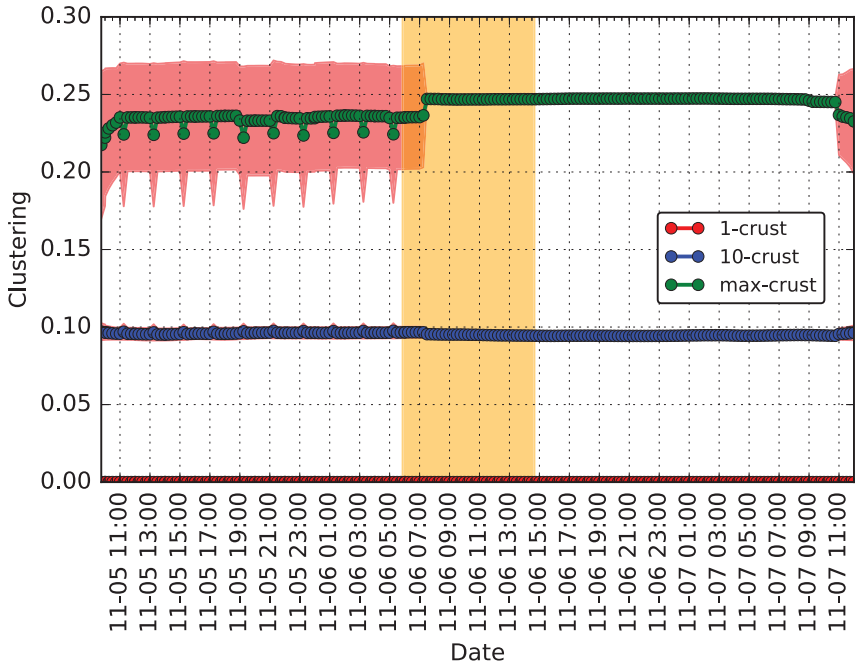

Figure 6.16: Clustering per crust India event.

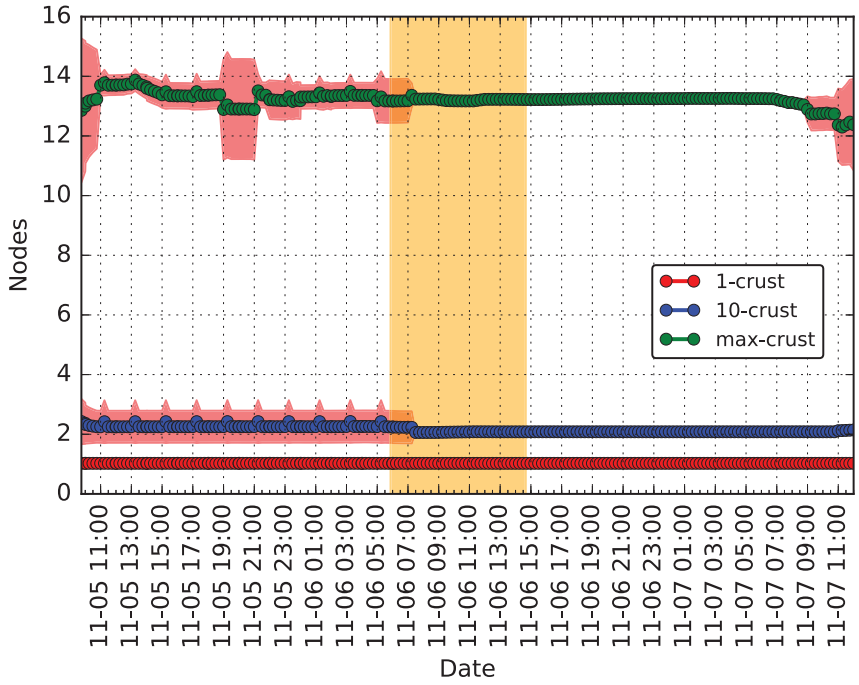

Figure 6.17: Average size components crust India event.

position, we have identified a method that is a complement to the current control- and data-plane anomaly detection approaches.

We are not proposing a product for technology transfer nor a real-time demonstration system. Instead this we present a promising methodological approach, identify the arguments of why it may be applicable, and then test the approach in three case studies.

To find the most useful properties to study control-plane anomalies, we used the k-shell decomposition of the AS-level graphs. Empirically, we noticed that both the core and the crust (for various k-levels) of these representations change, in some cases, more dramatically than in others. This is particularly of interest when considering the origin of the hijacks. Specifically, previous researchers have noted that the majority and more harmful attacks are orchestrated from peripheral ASes [Lad et al., 2007]. During the anomalous events, the incorrect routing information shifts the AS that is the source of the anomalous information from crust to core, as revealed in the k-shell decomposition.

We used centrality, average path length, and clustering properties in this dynamic analysis of the AS-level graph. We characterized the network before and after the anomalies (i.e., in the absence and presence of each anomaly). The properties of the decomposed crust and core graphs remain relatively constant without the anomaly. Then there is a sudden change in the crust graph immediately before each of the anomalies we examined were detected. Note however that these changes, although before the reported incidents, are not statistically significant.

Several factors can contribute to the difference in time at which we notice changes in the structural properties (with respect to the time announced by other researchers). First, there is a matter of sampling involved in the generation of the graphs. We use publicly available data that may have missing links; in particular, links between customers may be underrepresented [Oliveira et al., 2010].

It is worth noting that the presented characterization relies exclusively on the BGP data captured by the RouteViews project. That means that the characterized AS-level graph might not entirely represent the routing infrastructure at the moment of sampling. This is because route information does not necessarily take into account effective changes in the IP network ownership or commercial relationships between network operators. Thus, effective network routing changes

are not necessarily captured by the routers. This may cause a lack of precision when mapping the AS-level graph, and thus, the characterization of structural network properties derived from the graph. This may influence the accuracy of our construction of the graph from the AS-level topology. In addition, the study that we perform in this paper is based on the effect of a large-scale events. Therefore, it will be interesting to evaluate if the current approach also applies for other hijack events in which there are fewer number of IP prefixes compromised.

Another important factor to consider is the differentiation between changes in the topological properties when there is an incident and when there is not. The results of the research reported in this chapter illustrate that while some incidents can be detected, the method can not reliably distinguish the presence of an incident from its absence. Specifically, we noticed that even after the occurrence of incidents, there are changes in the properties of clustering coefficient and path length when we examined the graphs at different shells of the k-shell decomposition. Our explanation for this behavior is that the dynamics of BGP updates are busier independent if there are large-scale incidents of not. This explanation is supported by the findings in [Wang et al., 2002, Li et al., 2007]. This is important because our analysis in this chapter relies on processing BGP update's data. If there are many changes in the graphs (given a high number of updates) even in regular circumstances, then is going to be a major challenge to distinguish between benign and malicious dynamics.

Given that the analysis relies on assumptions concerning the generation of the observed data, the conclusions are also based on that data. Timing and scope are both issues in data compilation. It might be the case that the data is incomplete or delayed which would impinge our analysis. These considerations are important when trying to compare the effectiveness of the proposed approach with the traditional ones discussed above.

7 Bursty Announcements for Early Detection of BGP Routing Anomalies[1]

"The greatest challenge to any thinker is stating the problem in a way that will allow a solution."

— Bertrand Russell

7.1 Introduction

In this chapter, we propose a method for detecting large-scale Internet disruptions by analyzing the burstiness of BGP announcements. Our method tests the hypothesis that in order to perturb large parts of the Internet, before the incidents occur, the perpetrator will necessarily begin sending BGP announcement messages in a bursty manner, i.e., they occur in groups of relatively high frequency, followed by periods of infrequent activity. To do so, we propose an algorithm and show that for three large-scale incidents: the Indosat hijacking event in April 2014, the Telecom Malaysia leak in June 2015, and the Bharti Airtel Ltd, it is possible to provide early detection of them using measures from a technically diverse set of Internet's collector infrastructure.

7.2 Problem

The Internet, although extremely robust [Doyle et al., 2005], is notoriously vulnerable to attack by means of the Border Gateway Protocol (BGP) [Butler et al., 2010]. BGP exchange messages are assumed to be trustworthy. In other words, the reachability information shared between autonomous systems (ASes) is assumed to be correct without any verification. Despite the fact that

[1]The content of this chapter was submitted for review [Moriano et al., 2019b] at SIGCOMM in collaboration with Raquel Hill and L. Jean Camp. Pablo Moriano is the primary researcher and made all the analysis and figures therein.

the latest version of the BGP protocol was released in 2006 [Rekhter et al., 2006], there are no inherent protection mechanisms against participants advertising false routes.

In practice, BGP lacks authentication mechanisms not only for the announcement of the origin of IP prefixes but also the paths to that prefix. This leaves BGP vulnerable to unintended misconfiguration and malicious attacks [Goldberg, 2014]. The results of these disruptions include (i) traffic blackholing and (ii) interception. In traffic blackholing, the network traffic is dropped, never reaching its destination [Dyn Guest Blogs, 2008]. In traffic interception, the announcing AS reroutes traffic for the victim IP prefix and redirects it to the original origin AS after interception [Dyn Guest Blogs, 2013]. On this misdirected route, the traffic may be subject to eavesdropping [Arnbak and Goldberg, 2015], traffic analysis [Sun et al., 2015], or tampering [Shaw, 2013].

Well-known examples of BGP anomalies include the China Telecom hijack in 2010 [Hiran et al., 2013], the targeted interception of U.S. Internet traffic through Iceland and Belarus in 2013 [Peterson, 2013], and the large Indonesia ISP hijack in 2014 [Zmijewski, 2014, Toonk, 2014]. During the China incident, a routing update caused a large fraction of the world's Internet traffic (approximately 50,000 IP prefixes) to be redirected to China Telecom. This constitutes a very well-known example of a blackhole. A more recent incident was initiated by Indonesia's largest communication provider, Indosat. This incident was even larger than the China Telecom incident. The Indonesian ISP hijacked more than 320,000 routes. This means that Indosat laid claim to roughly two-thirds of the Internet for almost three hours. These were all identified only the *after* widespread diffusion of the incorrect routing information.

The current approaches to the challenges of routing anomalies rely on (i) cryptographic authentication or (ii) anomaly detection. Cryptographic protocols include the Resource Public Key Infrastructure (RPKI) [Lepinski and Kent, 2012] for origin authentication, and BGPSec [Lepinski and Sriram, 2017] which offers the ability to authenticate an entire path. These approaches are powerful, but there has not been widespread adoption [Gill et al., 2011]. This may be because of processor requirements, memory requirements, or a lack of incentive alignment [Hall et al., 2014]. Cryptographic solutions are also expensive. Perhaps more importantly, it has been shown that even

with their widespread adoption, it will be not possible to avoid the occurrence of route leaks, such as the Malaysian incident in 2015 [Toonk, 2015a, Madory, 2015b].

Anomaly detection approaches rely on measuring the control-plane (using BGP feeds) or the data-plane (exploring reachability of IP addresses in suspicious announced routes), or a combination of both. Anomaly detection does not require changes in the protocol itself. They primarily are used in detecting anomalies based on passive or active measurements in order to alert operators for mitigation and response [Shi et al., 2012, Khare et al., 2012, Toonk, Zhang et al., 2010]. Anomaly detection approaches are reactive because they identify harm after disruptive updates have polluted some detectable threshold of ASes with fake announcements.

Here we propose a detection method that aims to identify incipient incidents before diffusion and harm, by identifying a routing event as it emerges. Our goal is to identify events several hours prior to the state-of-the-art detection method, BGPmon [Toonk]. To do this, we use control-plane data collected by the RouteViews [Meyer, 2004] and served by BGPStream [Orsini et al., 2016]. The key observation in our anomaly detection method is that there are bursty BGP announcements before new routes are adopted by neighbor ASes. We characterize bursty announcements through statistical analysis of inter-arrival times. We conduct a case-based systematic analysis of the changes of inter-arrival times that are associated with three well-known anomalous events.

7.3 Methods

To provide early indicators of large-scale disruptions, we leverage the statistic-based anomaly detection method SRI NIDES used in the intrusion detection context [Javitz and Valdes, 1993]. Essentially, our method considers route announcements as signals with expected patterns of behavior and detects deviations from the expected patterns. Our focus is on inter-arrival times rather than the specific content of the announcements themselves. Claiming illegitimate ownership of a significant fraction of the Internet requires transmitting correspondingly bursty announcements causing large perturbations in the patterns of route announcements.

7.3.1 Data Sources

7.3.1.1 BGP Data

We compiled BGP updates (announcements and withdrawals) using BGPStream[2]. Update timestamp accuracy is one second. BGPStream provides an open-source software framework for the analysis of historical and real-time BGP data [Orsini et al., 2016]. To do so, BGPStream extracts data directly from route collectors. A route collector (collector, hereafter) is a host running a collector process. The collector emulates a router that establishes BGP peering sessions with BGP routers. These collection points are real—actual operating routers known as feeders. There are two popular projects running route collector processes, RouteViews [Meyer, 2004] and RIPE RIS [RIPE NCC, 2011].

At the time of this writing, RouteViews and RIPE RIS operate 22 and 23 collectors which peer with hundreds of feeders [Gregori et al., 2012]. We acknowledge that there are other sources of BGP data, including network operators, other route collector projects such as BGPmon [Yan et al., 2009] (from Colorado State University[3]), and Packet Clearing House [House]. However, previous research has shown that there is a considerable overlap between the measurements from RouteViews and RIPE RIS projects [Chen et al., 2009]. In addition, as pointed out in [Gregori et al., 2015], RouteViews provides the more complete view of the Internet in terms of IP prefix coverage. Therefore, we only collected BGP updates from RouteViews.

7.3.1.2 Routing Anomalous Events

Our data collection is based on a subset of BGP updates that cover the time before, during, and after selected incidents. We collected approximately seven days of observations around the start date of each of them. The purpose of collecting data over this time period is to be able to distinguish between regular and anomalous behavior. We consider these exceptional routing incidents because

[2]Available at https://bgpstream.caida.org/

[3]This refer to the free BGP monitoring service available at https://www.bgpmon.io/

of their impact to the Internet, the sheer number of prefixes, and the fact that these incidents have not previously received detailed academic analysis so that we could not know their patterns of diffusion in advance. Details about the incidents and their respective dates and times are listed below. Events are listed in chronological order. Note that these anomalous events have been studied and corroborated from different sources. To summarize:

An Indonesian ISP hijacks the world. On April 2, 2014, starting at 18:26 UTC, Indosat (one of the largest telecommunications providers in Indonesia) announced more than 320,000 IP prefixes belonging to other networks. Indosat announced roughly two-thirds of the entire Internet address space [Zmijewski, 2014, Toonk, 2014]. A large fraction of the hijacked prefixes belonged to Akamai, which is one of the larger Content Delivery Networks. This incident lasted approximately for 2.9 hours until 21:15 UTC. Traffic continued to be delivered; however, the path of the traffic was significantly altered.

Global collateral damage of the Telecom Malaysia leak. On June 12, 2015, starting at 08:43 UTC, Telecom Malaysia announced about 179,000 IP prefixes to Level 3 (the largest crossing AS) [Toonk, 2015a, Madory, 2015b]. Level 3 accepted these announcements and then propagated the routes to their peers and customers around the word. Because Telecom Malaysia is a customer of Level 3, the routes announced by Telecom Malaysia were identified as a preferred delivery route for Level 3. This event caused a significant packet loss and Internet service degradation around the world. Level 3 suffered a significant blackout from the Asia pacific region and the rest of the world. Note this was a leak, so the data were not delivered after being transmitted to Telecom Malaysia. This incident lasted approximately 2.7 hours. At around 10:40 UTC there were slowly observed improvements, and by 11:15 UTC the errors in the Routing Information Base (RIB) [Rekhter et al., 2006] began to be resolved.

Large scale BGP hijack in India. On November 6, 2015, starting at 05:52 UTC, Bharti Airtel Ltd., claimed the ownership of about 16,123 IP prefixes. These addresses corresponded to more than two thousand unique ASes [Toonk, 2015b, Murphy, 2015]. This event became widespread because two large ASes (e.g., Cogent Communications and GlobeNet Cabos Submarinos S.A.) ac-

cepted and propagated these routes to their peers and customers. Legitimate owners of the prefixes included Akamai, Tata Communications, and Apple Inc. This incident lasted approximately 8.9 hours until 14:40 UTC.

7.3.2 Burstiness of Announcements

Let $X_{A\to B} = \{X_{A\to B}(t)\}, t = 0, 1, \ldots, N$ be a time series of time-stamped announcements sent by AS A and received by collector B. Let $\tau_{A\to B}$ be a random variable that represents the time interval between consecutive announcements so that $\tau_{A\to B}$ takes values in $\{X_{A\to B}(1) - X_{A\to B}(0), X_{A\to B}(2) - X_{A\to B}(1), \ldots, X_{A\to B}(N) - X_{A\to B}(N-1)\}$. Burstiness refers to the tendency of certain events to occur in groups of relatively high frequency, i.e., short inter-arrival time intervals, followed by periods of relatively infrequent events [Harang and Kott, 2017]. Mathematically, it can be characterized by analyzing the inter-arrival time distribution $P(\tau_{A\to B})$. As was proposed in [Goh and Barabási, 2008], the inter-arrival distribution can be characterized by a burstiness factor defined by $B = \frac{\sigma-\mu}{\sigma+\mu}$. Here σ and μ denote the standard deviation and mean of the inter-arrival time distribution. Note that the burstiness has a value of -1 for $\sigma = 0$, which means regular time intervals. It has a value of 0 for $\sigma = \mu$ in the case of random time intervals. Finally, it has a value of 1 for $\sigma \to \infty$ and a finite μ in the case of a highly bursty time series of announcements.

7.3.3 Detection Method

We leverage the measure of inter-arrival times as received by the collectors to compute a measure of intensity based on the burstiness of announcements. This measure was originally used in the context of intrusion detection in [Javitz and Valdes, 1993]. Let $Q_{A\to B}$ be the number of announcements sent by AS A and received by collector B exponentially weighted. This means that more current announcements have a greater impact in its computation, i.e., short inter-arrival times. The value of $Q_{A\to B}$ is computed using the recursive formula

$$Q_{A\to B}(t) = 1 + 2^{-r\Delta} Q_{A\to B}(t-1). \tag{7.1}$$

Here r is the decay factor and $\Delta = X_{A\to B}(t) - X_{A\to B}(t-1)$ is the inter-arrival time between consecutive announcements. The decay factor r determines the half-life of $Q_{A\to B}(t)$. Large values of r imply that the value of $Q_{A\to B}(t)$ is more influenced by more recent announcements. Smaller values of r imply that the value of $Q_{A\to B}(t)$ will be more heavily influenced by announcements in the distant past. Detection focuses on identifying observations in the time series of $Q_{A\to B}$, for which its value exceeds a threshold that is a function of the mean and standard deviation. We use the moving average and moving standard deviation as the mean and standard deviation estimators respectively. The parameter ω is the window length in the moving average model. The parameter δ controls how many standard deviations are considered to report an event. The complete pseudocode for the detection algorithm can be found in below.

Algorithm 5 Event-Detection ($X_{A\to B}$, r, ω, δ)

```
1: Q_{A→B} ← {}                                          ▷ Number of announcements
2: for t in X_{A→B} do
3:     Q_{A→B} ← Q_{A→B} ∪ {1 + 2^{-rΔ} Q_{A→B}(t-1)} using eq. (7.1)
4: end for
5: Ψ ← moving average(Q_{A→B}, ω)
6: Σ ← moving std(Q_{A→B}, ω)
7: Ê ← {}
8: for t in {0, 1, ..., N} do
9:     if Q_{A→B}(t) >= (Ψ(t) + δΣ(t)) then
10:        Ê ← Ê ∪ {t}
11:    end if
12: end for
13: return Ê
```

7.4 Results

In this section, we analyze the previously described BGP incidents. We analyzed the views from several data collectors at various locations around the world. Table 7.1 shows the geographical location and the date of the first dump of the collectors used in this study[4]. We analyzed BGP announcements and withdrawals, but the withdrawals did not effect our results, perhaps in part because the volume of withdrawals is significantly less [Wang et al., 2002, Lad et al., 2003, Deshpande et al., 2004]. Our results presented here include only announcements.

[4]Collectors' location and date of the first dump were obtained from RouteViews and BGPStream respectively.

Table 7.1: Geographical location and date of first dump of collectors. Collectors are ordered in alphabetical order.

Collector name	Location	First dump
route-views.chicago	Chicago, IL, US	2016-06-28 12:00
route-views.eqix	Ashburn, VA, US	2004-05-17 13:59
route-views.isc	Palo Alto, CA, US	2003-11-27 02:00
route-views.jinx	Johannesburg, ZA	2012-07-10 00:00
route-views.kixp	Nairobi, KE	2005-10-07 15:44
route-views.linx	London, GB	2004-03-16 13:45
route-views.nwax	Portland, OR, US	2014-03-20 20:52
route-views.perth	Perth, AU	2012-11-15 21:48
route-views.saopaulo	Sao Paulo, BR	2011-03-17 16:19
route-views.sfmix	San Francisco, CA, US	2015-04-14 20:00
route-views.sg	Singapore, SG	2014-06-04 15:44
route-views.soxrs	Belgrade, RS	2014-01-01 00:00
route-views.sydney	Sydney, AU	2010-08-14 02:00
route-views.telxatl	Atlanta, GA, US	2012-02-02 22:46
route-views.wide	Tokyo, JP	2003-07-01 21:29
route-views2	Eugene, OR, US	2001-10-26 16:48
route-views3	Eugene, OR, US	2013-11-25 10:00
route-views4	Eugene, OR, US	2008-11-28 09:53
route-views6	Eugene, OR, US	2003-05-03 12:29

We conduct four different but complementary analyses. First, we monitor the dynamic behavior of the number of feeders peering with each collector. To this end, we perform a longitudinal analysis that spans over 17 years to quantify the trends in contribution of feeders to the collectors. This analysis shows the time when some collectors started attracting or being disconnected from some feeders—we relied on the number of routers for this. A large number of feeders produces a robust view of Internet activity. This indicates the time-frame when collectors could construct a representative view of Internet activity. This analysis is complementary to the works in [Gregori et al., 2012, 2015] that focused on a single month of observations and other works [Wang et al., 2002, Li et al., 2007, Deshpande et al., 2004, Zhang et al., 2004] that analyze a stream of BGP updates as seen from a single collector.

Second, we show how each of the large-scale incidents is perceived from the point of view of the different collectors. To do so, we measure the number of announcements received by the collectors before, during, and after the incident. We study how announcements vary based on the

number of feeders into a particular collector. We found that the incident can be viewed more clearly from those collectors with more feeders. This is particularly true for collectors in North America and Europe. Some collectors are unable to detect the incident because, with a small number of feeders, the collector is unable to construct a robust view of the Internet. This is a result of the low number of feeders peering with the collectors, as exemplified in the case of several African countries [Gregori et al., 2017].

Third, we analyze the inter-arrival times of announcements at the collectors. We characterize these with a measure of burstniess used previously for studying human dynamics in [Goh and Barabási, 2008, Barabási, 2005]. This allows us to quantify the burstiness of announcements before, during, and after the incidents. We show that ASes involved in the reported incidents exhibit a statistically significant change in the inter-arrival pattern of their BGP announcements at the collectors. We show that for early detection of BGP incidents, the volume of messages is not enough for incident detection. In contrast, the burstiness of the announcements sent by ASes and seen for a specific collector is a complementary discriminator to the volume of announcements and helps identify anomalous behavior. More importantly, we show that changes in burstiness occur several hours before the incidents were reported in practice.

Fourth, we propose a method for detecting anomalous announcements based on quantifying the burstiness of announcements that are received by the collectors. This allows us to characterize the distinguishing features that occur before the incidents. Given these distinguishing features, we introduce a detection algorithm and evaluate its effectiveness using real-stream data obtained from collectors during the incidents.

7.4.1 Feeder Contribution Analysis

To quantify the number of feeders peering with each collector, we parse RIB dumps of each collector and count the number of unique peering routers. We counted the number of unique routers on the first day of each month at noon for 195 consecutive months from October 2001 to December 2017. This allowed us to study the evolution of the number of feeders per collector.

Figure 7.1 shows the time series of the number of feeders based on the number of routers. We see that, in general, that there is an increase in feeders over time. In particular, route-views.linx, route-views.saopaulo, and route-views4 report the highest number of feeders by the end of the observation period. Conversely, route-views.kixp, route-views.soxrs, and route-views.wide show a significantly lower number of feeders, confirming the recent findings by Gregori et al. in [Gregori et al., 2017].

There are also some collectors with fluctuations in this time period. In particular, the number of peering routers seems to decrease by mid 2016 for route-views.saopaulo and route-views4. Despite this, even more importantly, route-views4 has the highest feeder count. This is important to our work because the capacity to detect the incidents depends on collectors having a robust view of the Internet so that they may construct an accurate baseline. Therefore, collectors need to have a minimum number of feeders to be able to see the incident.

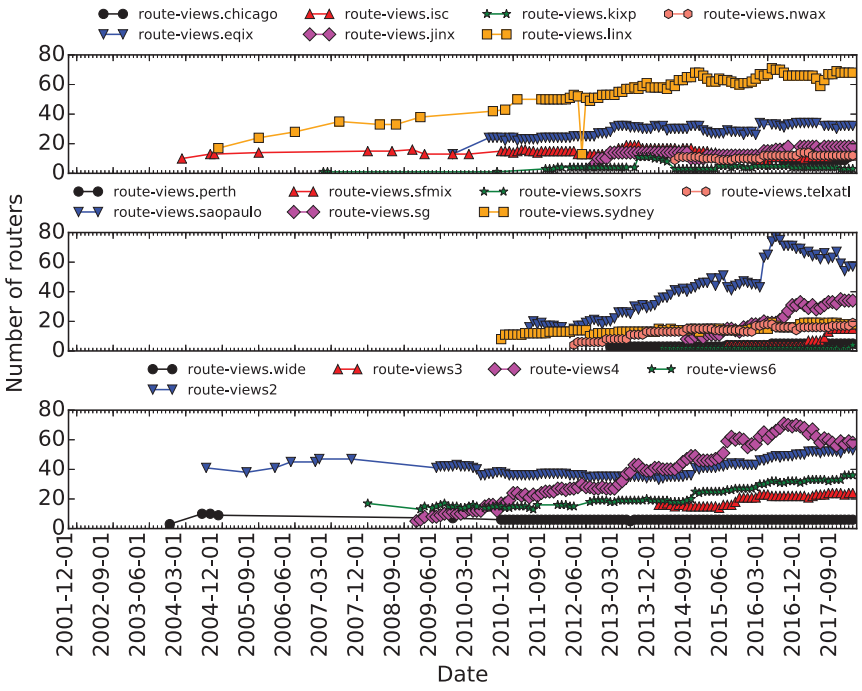

Figure 7.1: Time series of the number of routers peering with collectors. Collectors are ordered in alphabetical order. Major ticks correspond to nine-month intervals while minor ticks correspond to one-month intervals.

7.4.2 Collectors' Disruption Perception

We show observations during a seven-day period around the occurrence of the incidents. These are highlighted in the plot between the two vertical dashed lines as reported by the state-of-the-art BGP anomaly detection system BGPmon [Toonk]. We ranked the collectors in decreasing order by the number of feeders. We show the view from the four collectors with the highest number of feeders. Note that the events are observable by analyzing the view of the four collectors before the incidents were reported by leading third-party services including as Oracle Dyn and BGPmon respectively [Zmijewski, 2014, Toonk, 2014].For the remaining collectors please refer to Appendix B.1.

Indosat incident. Figure 7.2 shows the number of announcements received from the AS responsible for the incident, i.e., AS 4761. This incident is perceived differently at each collector. The incident is almost unnoticeable for collectors with a low number of feeders (route-views.soxrs, route-views.perth, and route-views.kixp with 2, 3 and 3 feeders respectively). For the other collectors, two things happen. First, there is a significant increase in the number of received announcements. This increase is almost four orders of magnitude in the majority of the cases. Second, the frequency at which the announcements are received is higher than other announcements that are not close to the occurrence of the incident, i.e., around the highlighted region. This last observation implies shorter inter-arrival times in the proximity of the incident. It is worth noting that for the collectors which received announcements, this striking behavior is perceived almost four hours before the incident.

Telecom Malaysia incident. Figure 7.3 shows the number of announcements received by every collector, recall the originator is AS 4788. Some collectors observe an increase in burstiness in announcements received prior to the incident. The number of announcements increases up to four orders of magnitude. Even more importantly, these announcements occur highly intermittently and frequently. As in the case of the Indonesia incident, some collectors did not see the behavior that we are describing, e.g., route-views.sorxs and route-views.kixp, with two and four feeders respectively. Note that this pattern of behavior occurs almost three hours before the incident was

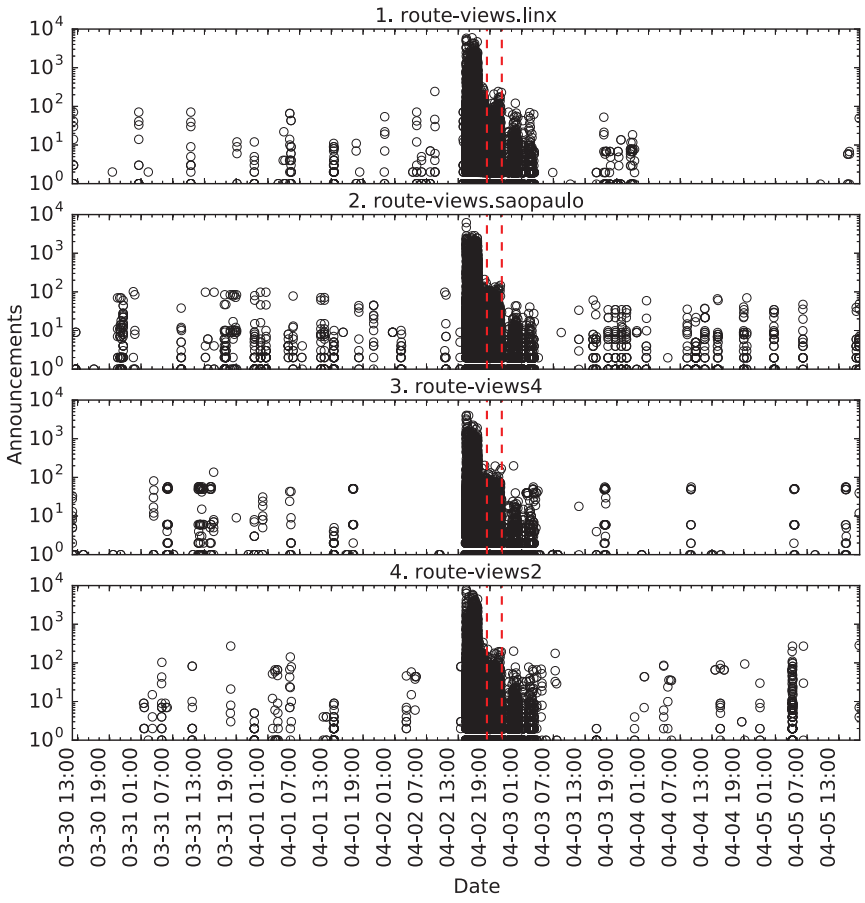

Figure 7.2: Time series of the number of announcements from AS 4761 that collectors received before, during, and after the Indosat incident in 2014 for the top four collectors. Major ticks correspond to six-hour intervals while minor ticks correspond to two-hour intervals.

detected.

Bharti Airtel Ltd. incident. Figure 7.4 shows the number of AS 9498 announcements received by the collectors. As with Indonesia and Malaysia, the incident is seen more clearly from some collectors than from others, and from some not at all. For collectors where the incident would have been detectable, the number of announcements increases up to five orders of magnitude. Note that the the bursty behavior of the announcements right before the incident is less intense than in the previous cases. Again, this burstiness is clearly noticeable several minutes before the incident was detected on the network.

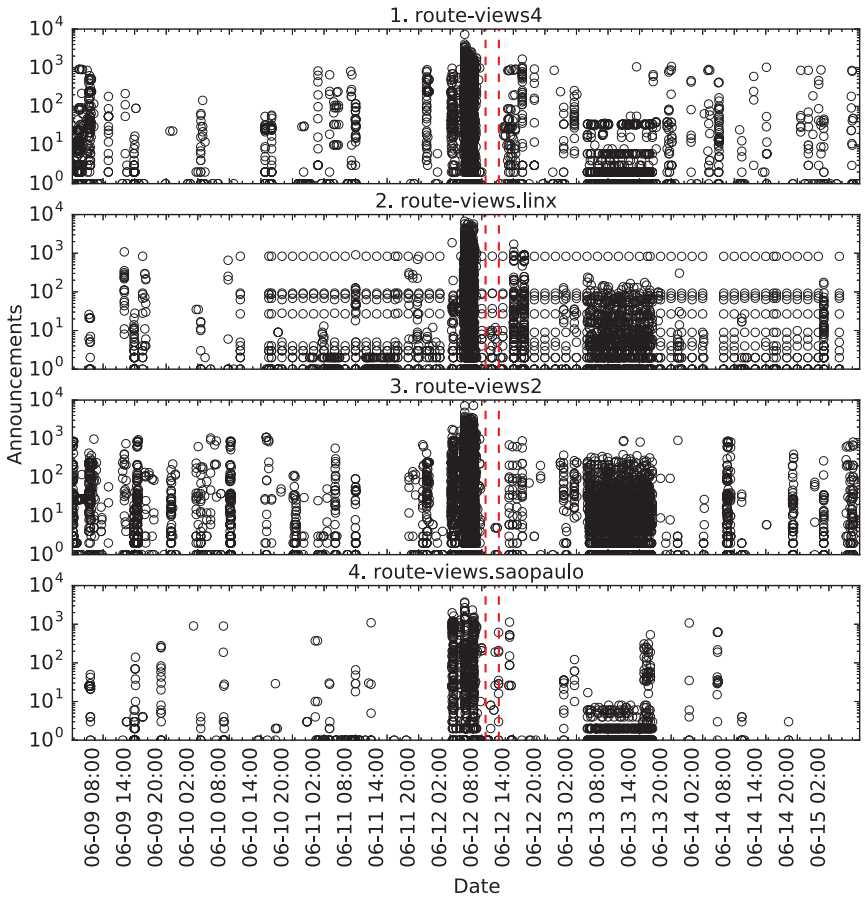

Figure 7.3: Time series of the number of announcements from AS 4788 that collectors received before, during, and after the Telecom Malaysia incident in 2015.

7.4.3 Inter-Arrival Time Analysis

There is both a significant increase in the number of arrivals of announcements before the incident (see Section 7.4.2) and a dramatic increase in the frequency at which these announcements are received by the collectors. The following analysis reveals that the inter-arrival time of announcements as seen by the collectors exhibits a significant degree of burstiness. To ground the results, we first analyze the joint distribution of activity of each AS based on the burstiness (horizontal axis) and the number of announcements (vertical axis) during one full day of measurements around the incident. Collectors are ranked in decreasing order by number of feeders. We provide details for the top four collectors in this work. To assure an accurate assessment of burstiness, we only consider ASes that sent more than 100 announcements during this time interval [Kim and Jo, 2016].

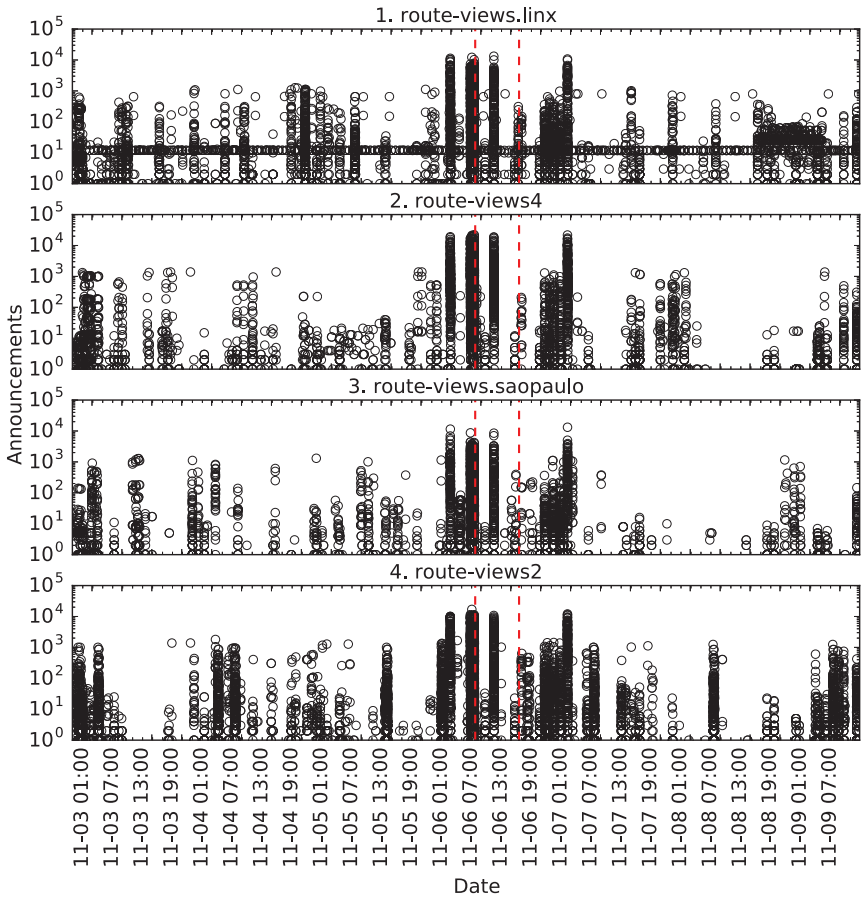

Figure 7.4: Time series of the number of announcements from AS 9498 that collectors received before, during, and after the Bharti Airtel Ltd. incident in 2015.

We marked with "squares" the ASNs of the top five ASes based on CAIDA's customer cone size ranking [cai, 2018], i.e., AS 3356, AS 1299, AS 174, AS 2914, AS 3257. They provide a baseline for comparison. We mark with a "star" the ASN that was responsible for the incident. The dark cells indicate a high concentration of ASes with a characteristic burstiness and number of announcements.

Second, we test if the apparent effect is real or is it due to chance. In particular, we apply a Monte Carlo test in which the null hypothesis is that ASes send announcements in a bursty manner even during times where there is no evidence of a BGP incidents. For this analysis, we collected time series of announcements over a full day of observations where no BGP incidents have been detected. One hundred of these random time series were compiled for each collector for the top

five ASes (again based on CAIDA's customer cone size ranking) and the AS involved in each incident. In each of these 100 time series, we compute the ASes associated burstiness. Here we provide the results for the top four collectors based on the number of feeders, again with data and details for the other collectors available upon request. For the remaining collectors please refer to Appendix B.2.

Indosat incident. Figure 7.5 shows that during the incident, most of the ASes have a burstiness that is below 0.90 (the 93th percentile) and produce fewer than 10^6 announcements (the 97th percentile). There are dashed lines on these percentiles. Note that the AS represented by the star (i.e., Indosat) has the highest burstiness. Note also that those ASes in the second quadrant (with more than 10^6 announcements) have a considerable number of announcements but lower burstiness (i.e., AS 27738, AS 9829, AS 53062, AS 36998, AS 29571). These ASes appear consistently among the different collectors but were not reported to be involved in the incident. Conversely, those ASes in the fourth quadrant show high burstiness, but the number of announcements is not significant (i.e., AS 7629, AS 61125, AS 9497, AS 132045). We found that those ASes are not neighbors of Indosat (corroborated through [cai, 2018]) nor involved in the incident. This empirical finding reveals that although the volume of announcements increases for different ASes during the incident, the actual AS involved in the incident has a distinct burstiness pattern that starts several hours before the incident was reported. This observation is complementary to the works in [Lad et al., 2003, Deshpande et al., 2004] in which a significant increase in the volume of announcements is used as a detection signature, as well as illustrating the benefit of including a measure of burstiness.

Figure 7.6 shows notched box plots comparing the burstiness calculated from different collectors for the baseline ASes and the AS involved in the incident (the last one) under the null hypothesis. Notched box plots have a contraction around the median whose height is statistically important. When notches of the boxes overlap, there is not a statistically significant difference between the medians. In this case, these plots illustrate that the burstiness of each of the ASes under study are not significantly different when there is no incident. However, the observation highlighted with the cross corresponds to the test statistic for the observations derived during the

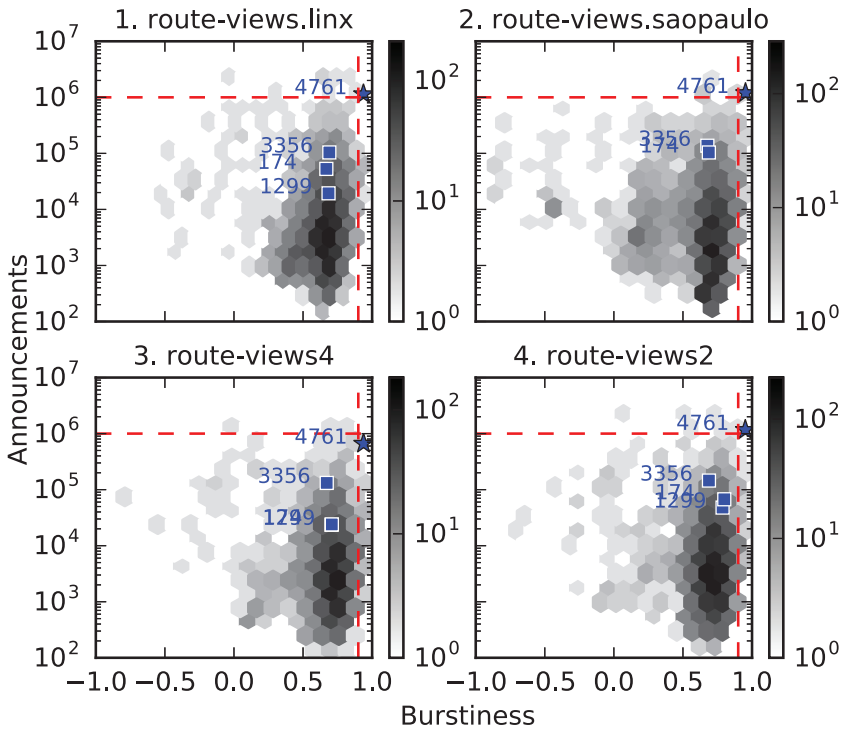

Figure 7.5: Joint distribution based on the the burstiness (horizontal axis) and number of announcements (vertical axis) during one day interval around the Indosat incident.

interval of the incident. As can be seen, for collectors receiving announcements from the AS involved in the incident, this observation lays outside the region of statistical indistinguishably. This suggests that the burstiness during the incident is statistically significant different, and it is unlikely that such values would be observed under random conditions. This argument reinforces the argument that the volume of announcements is a necessary but not sufficient feature for early detection of large-scale BGP incidents (see Fig. 7.5). High burstiness is a distinctive feature in these incidents.

Telecom Malaysia incident. Figure 7.7 shows that the AS involved in the incident has a distinct characterization in the distribution, i.e., AS 4788. It has both high burstiness and number of announcements. Most of the ASes have a burstiness that is below 0.90 (the 99th percentile) and produce less than 10^6 announcements (the 99th percentile). Note that there are ASes that sent a high number of announcements and do not have high burstiness compared to Telecom Malasya (those in the second quadrant), e.g., AS 9892, AS 9829, AS 9583. These ASes were not involved

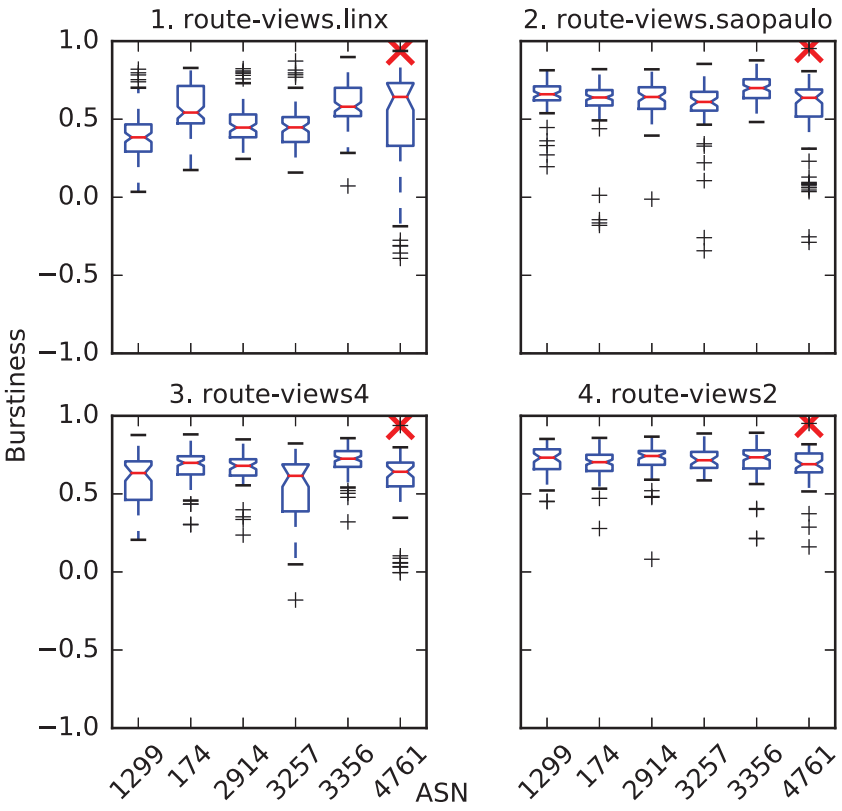

Figure 7.6: Monte Carlo test for burstiness. Last column corresponds to the observations of the AS responsible for the incident, AS 4761. The test statistic, the burstiness observed during the interval of the attack, is marked with a cross.

with the incident nor are they neighbors of Telecom Malaysia. Conversely, the ASes in the fourth quadrant have higher burstiness but fewer announcements compared to Telecom Malaysia, e.g., AS 28681, AS 10208, AS 8402, AS 45209. These are not neighbors of Telecom Malaysia but there is no evidence of malicious updates coming from them.

Figure 7.8 shows the distribution of burstiness computed over 100 samples of random one day intervals. The burstiness of Telecom Malaysia is highlighted with the cross. This figure shows that the burstiness of the AS that was involved in the incident is statistically significantly larger when compared to its own normal behavior (e.g., baseline and null comparisons).

Bharti Airtel Ltd. incident. Figure 7.9 shows that the burstiness of AS 9498 is not as high as the burstiness of the perpetrators of the incidents in Indonesia and Malaysia. Here most of the ASes have a burstiness that is below 0.90 (the 99th percentile) and produce fewer than 10^6

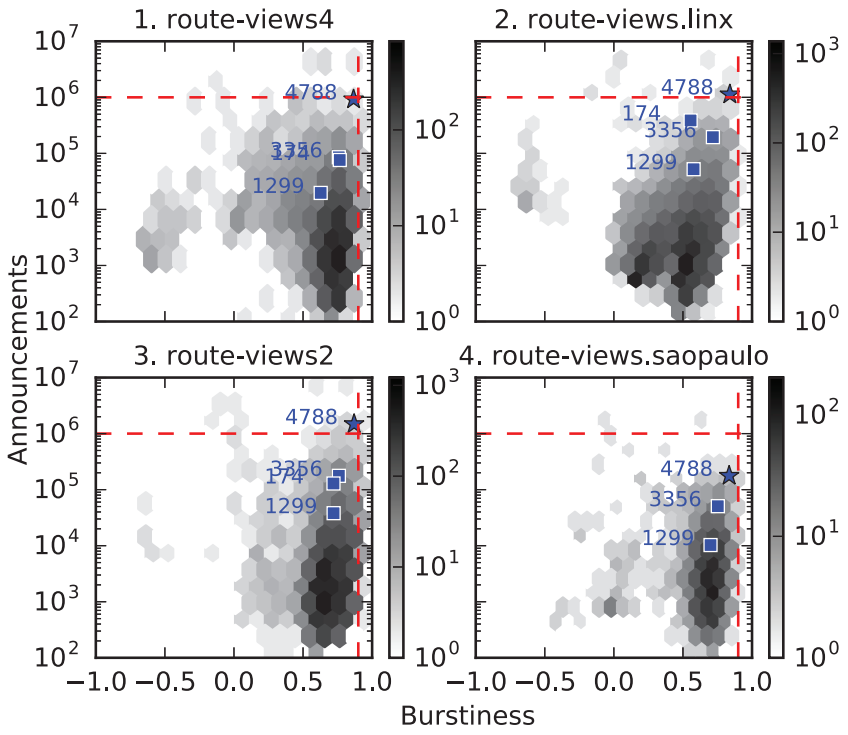

Figure 7.7: Joint distribution based on the total number of announcements and their burstiness during one day interval around the Telecom Malaysia incident.

announcements (the 99th percentile). However, for the majority of the collectors, Bharti Airtel sent a large number of announcements that placed it in the second quadrant. However, there are other ASes in the second quadrant, e.g., AS 262949, AS 9829, AS 36408, AS 28573, AS 21669. From these, AS 9829 and AS 36408 are customers of AS 9498. Conversely, we find that AS 394104, AS 11139, AS 133722, and AS 42040 have high burstiness and relatively fewer announcements, but they are not neighbors of AS 9498 nor involved with the incident.

Figure 7.10 shows the Monte Carlo test for AS 9498 and the top ranked five ASes. We observe that the burstiness of AS 9498 is at the boundary of the distribution but not as significant as in the Indonesia and Malaysia cases. This reflects the fact that for this incident, announcements were less bursty. They are significant when compared with the normal behavior of AS 9498.

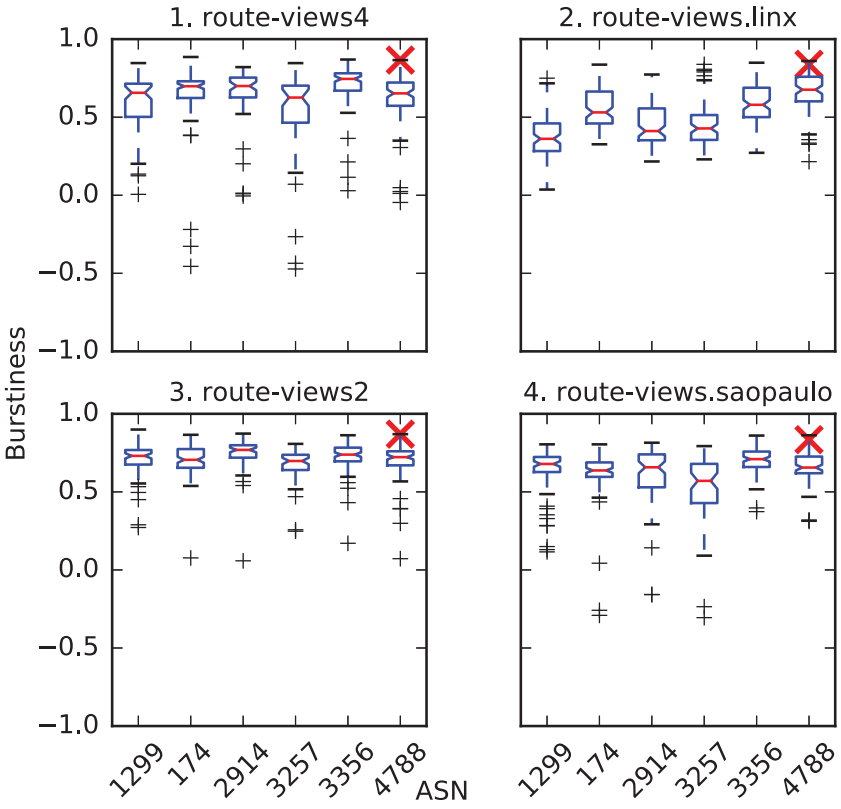

Figure 7.8: Monte Carlo test for burstiness. The last column corresponds to the observations of the AS responsible for the incident, AS 4788.

7.4.4 Anomaly Detection

The main idea of our anomaly detection method relies on profiling the expected behavior of a signal and then detecting deviations from the expected pattern. To do so, we rely on the measure of burstiness of announcements as perceived by the collectors. Analyzing the volume of announcements can be misleading, and adding the measure of burstiness has two advantages. A high volume of announcements may be caused by BGP session resets and other vendor specific behaviors [Wang et al., 2002]. It enables earlier detection of anomalies and decreases the number of candidates to be examined as potential anomalies (e.g., quadrant two in Figs. 7.5, 7.7, 7.9).

To evaluate burstiness, we compute the time series $Q_{A\to B}$ for each incident, based on equation (7.1). The solid line represents the value of $Q_{A\to B}(t)$ for each arriving unique announcement message at time t. In accordance with previous studies in [Zhang et al., 2004, Labovitz et al.,

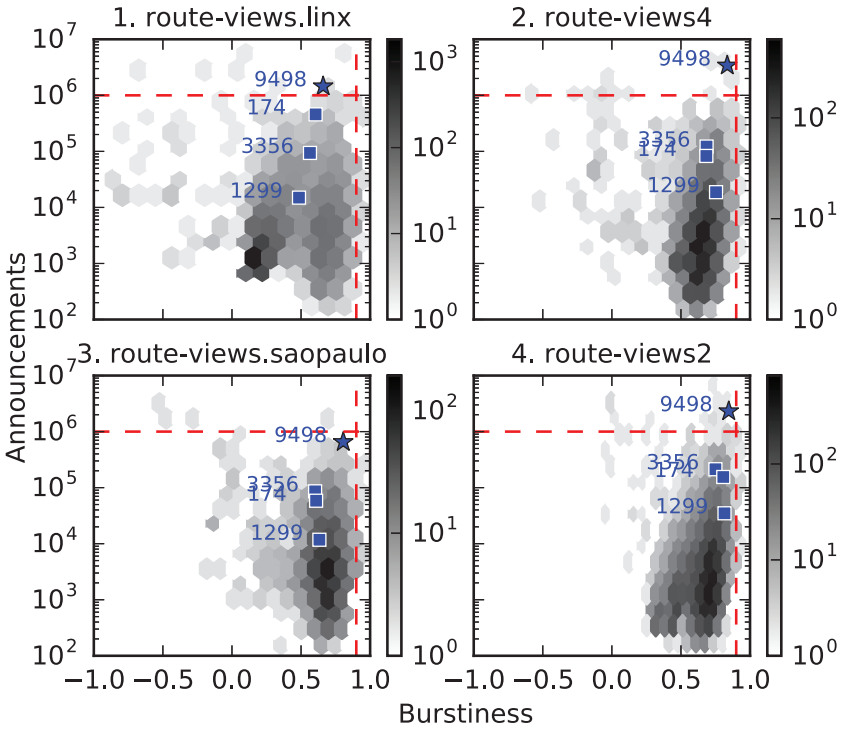

Figure 7.9: Joint distribution based on the total number of announcements and their burstiness during the one day interval around the Bharti Airtel Ltd. incident.

2000], we verify that most of inter-arrival times of announcements are less than 300 seconds (the 99th percentile for most of the collectors). We then use 300 seconds as the half-life value to capture most of routing dynamics. Then the decay factor is set to be $r = 1/300$. Each horizontal gray band represents one standard deviation from the moving average using the same window length. We use $\omega = 20$ as the estimator for the window length because it is the lowest value that reduces the mean square error between the empirical observations and the moving average. The darkness of the bands indicates the distance from the means based on Algorithm 2. Observations that lay more than two standard deviations away from the moving average are marked with stars, i.e., we use $\delta = 2$. Note that the values of r, ω, and δ may be tuned for detection purposes. We do that here for the same collectors as in the previous analysis. For the remaining collectors please refer to Appendix B.3.

Indosat incident. Figure 7.11 shows that almost four hours before the event was reported, $Q_{4761 \to B}$ is more than two standard deviations away from the moving average. Interestingly, the data from

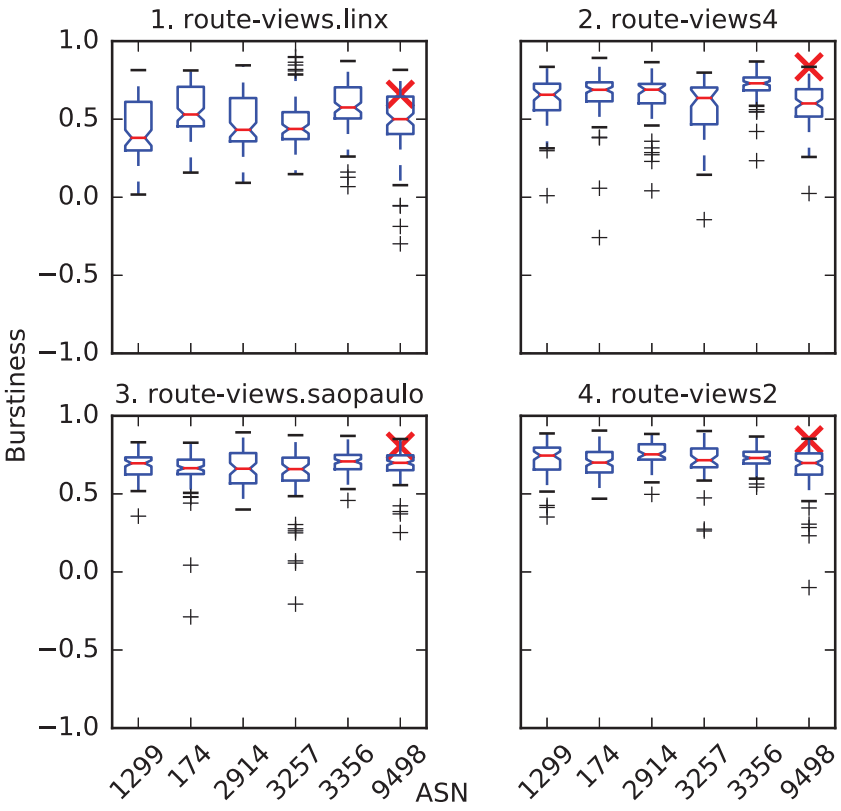

Figure 7.10: Monte Carlo test for burstiness. The last column corresponds to the observations of the AS responsible for the incident, i.e., AS 9498.

route-views.linx— the collector with the highest number of feeders— is the first to deviate from the mean, specifically 3 h 43 min and 2 seconds before the earliest detection of the incident. The deviations of the other collectors are later but still hours before the incident was actually detected. Note these outliers do not show up at other dates or times of the time series.

Telecom Malaysia incident. Figure 7.12 shows the time series of $Q_{A\to B}$ for four collectors. The value of $Q_{4788\to B}$ is more than two standard deviations almost four hours before the incident was reported by BGPmon. Here the collector with the more anticipated observation is route-views4, with anomalous readings clear 3 h 51 min and 2 seconds before detection. Note also that route-views.saopaulo reports no outliers, meaning that the perceived burstiness is not as high as for the other collectors (see Fig. 7.7).

Bharti Airtel Ltd. incident. Figure 7.13 shows that collectors observe anomalies in advance of the detection of the incident. However, route-views2—the collector that received the burstiest

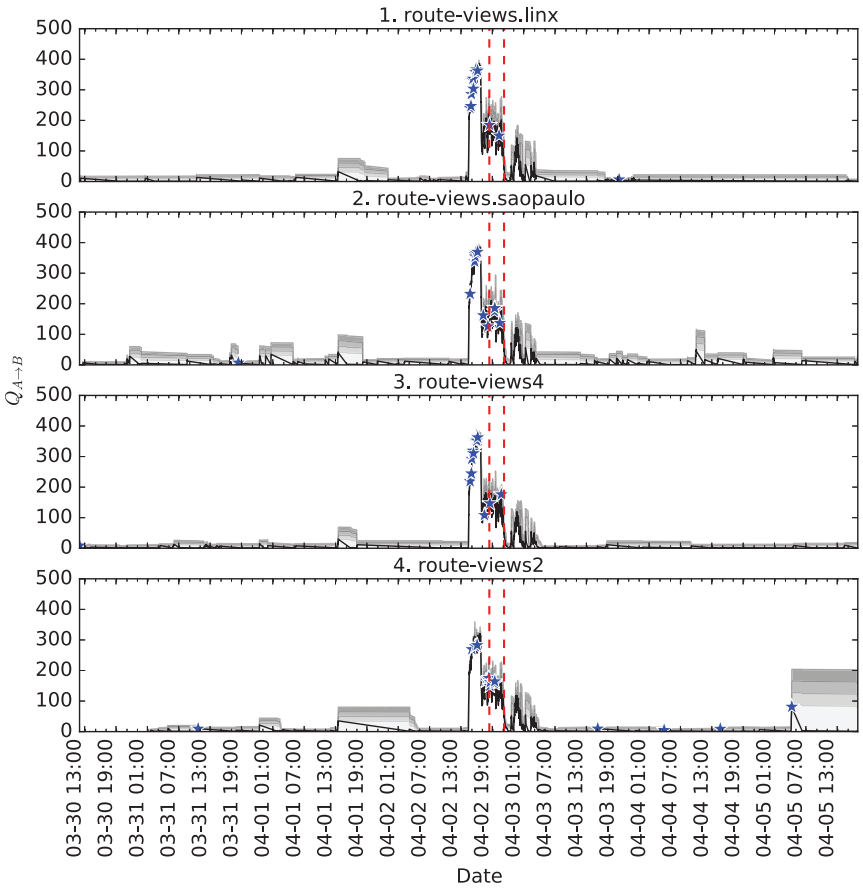

Figure 7.11: $Q_{4761\to B}$ time series for the Indosat incident.

signal according to Fig. 7.9—observes these outliers only 25 min before. The lower impact may be correlated with lessor potential for advance notice, but there is no data to assert this as a conclusion.

7.5 Conclusion

Routing anomalies caused by both misconfigurations and malicious intent have tested the resilience of Internet core protocols [Moriano et al., 2017c]. Here, we propose an anomaly detection method and show that it would identify three large scale anomalies significantly in advance when compared with the state-of-the-art method [Toonk]. To do so, we analyze inter-arrival times of BGP announcements leveraging the RouteViews collector infrastructure. We found that the burstiness, along with the volume of announcements, has the potential to provide early warnings of routing

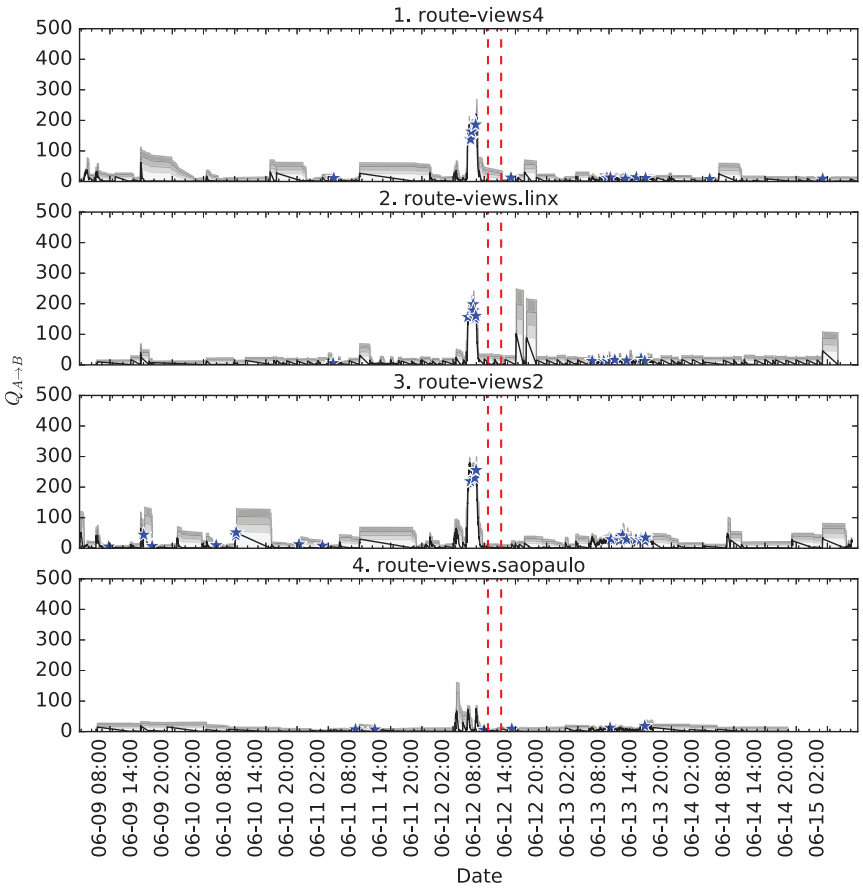

Figure 7.12: $Q_{4788\to B}$ time series for the Telecom Malaysia incident.

anomalies before they are evident using traditional control-plane and data-plane approaches. We believe that the proposed method is a complement to current anomaly detection approaches.

To validate the effectiveness of the proposed method, we conducted analysis for three cases of large-scale routing anomalies. We have evaluated the statistical significance of announcement burstiness, before, during, and after the events. We found that the perpetrators of the incidents have statistically significant bursty patterns that are visible from the collectors several hours before the incidents were reported by others. We analyze the same features under the null case (of no incidents) and corroborate that the bursty behavior is characteristic of announcements sent prior the detection of the incidents. By relying on this key observation, we propose an algorithm to identify when there is an incipient anomalous incident. The data and scripts used in this research will be made available for reproducibility purposes.

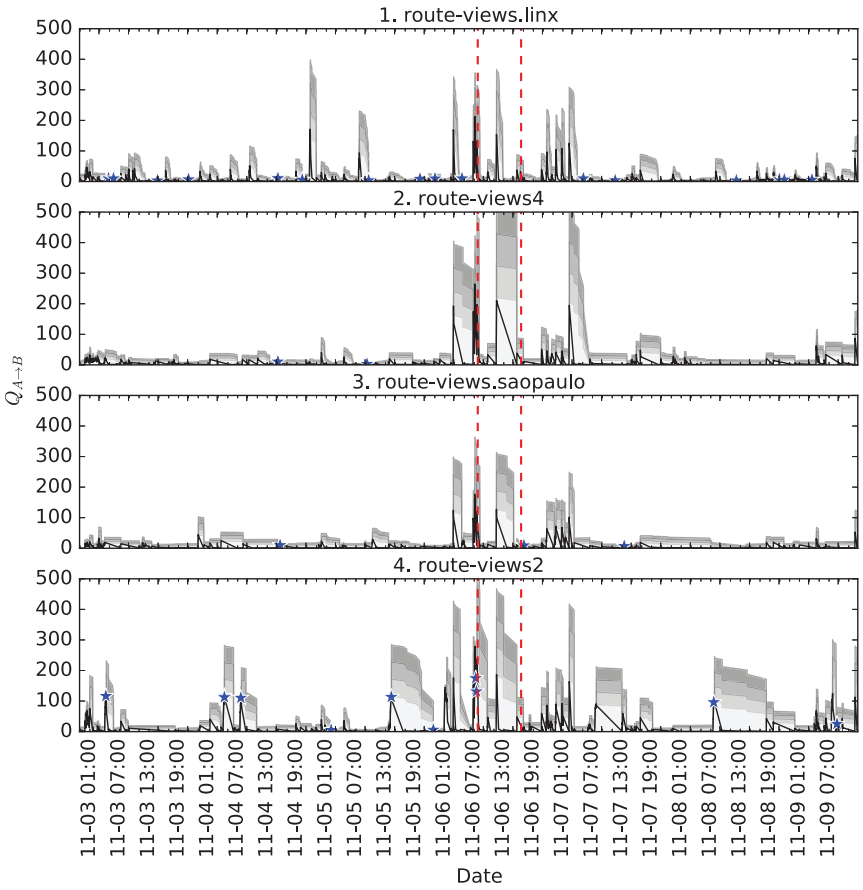

Figure 7.13: $Q_{9498\to B}$ time series for the Bharti Airtel Ltd. incident.

The proposed method would be effective against hijacks, route leaks, and other misconfigurations. Having noted the potential for our approach, we are also aware of some limitations of our proposed work.

Real-time data availability: Our analysis is based on BGP announcements received by RouteViews collectors. Only a subset of these collectors support real-time monitoring through BGPmon[5]. The RouteViews data used in our analysis relies on BGPStream, which has an access delay of approximately 20 min [Sermpezis et al., 2018a]. One option for further research is to run these experiments with a reduced number of current real-time RouteViews collectors through BGPmon. In addition, RIPE RIS provides an API to access real-time BGP updates for a limited number of

[5]Here BGPmon refers to the free monitoring service develop by Colorado State University available at https://www.bgpmon.io/

collectors. Through sharing our scripts, we hope that individual collectors could implement this approach and report the results in the future.

Feeder contribution: Our method treats each router contribution as equivalent. In fact, they vary significantly in terms of IP space coverage as shown in previous research [Gregori et al., 2012, 2015].

Unknown efficacy for subtle attacks: We evaluated our method for large-scale high-impact routing incidents, both hijacks and route leaks. However, we do not investigate other attack configurations that are used for more subtle attacks. This analysis excludes incidents such as when U.S. Internet traffic was rerouted through Iceland and Belarus in 2013 [Peterson, 2013] and routing attacks on cryptocurrencies [Apostolaki et al., 2017].

Focus on early detection but not mitigation: We propose an anomaly detection method that allows early identification of BGP large-scale incidents. To do so, the effectiveness of our proof-of-concept is evaluated based on its ability to detect incidents before state-of-the-art detection methods. Yet we do not discuss mitigation strategies once the events are detected, e.g., prefix deaggregation [Lutu et al., 2012]. Of course, these mitigation strategies can be implemented on top of our proposed method to avoid wide diffusion of route misinformation.

Burstiness is a complement to volume measurements: Burstiness needs to be measured simultaneously with the volume of announcements. As confirmed by the results for the Indosat incident, burstiness and volume of announcements must be combined to reduce false positives and to provide early detection of the incidents.

Overhead in route collectors: Route collectors are instruments used for measurement in our proof-of-concept. For implementation purposes, the detection method might most effectively be implemented at the collectors. We do not expect these processes to impose significant overhead on the collectors, but we have no actual performance measurements.

8 Conclusions

"Education isn't something you can finish."

— Isaac Asimov

As stated in Chapter 1, this dissertation argues that by developing data-driven and network science methods, we can obtain a more comprehensive understanding of the emergence of anomalies in longitudinal security data. We created an innovative approach using previously developed methods from other domains to identify anomalies in a variety of security problems. Here we review those findings and provide some final remarks.

8.1 Summary of Contributions

Chapter 2 provides a general overview of graph-based anomaly detection and its application to computer security problems. We focused on BGP and insider threat anomaly detection. This was presented to provide the necessary grounding for the subsequent analysis in the dissertation. The following sections summarize the contributions of each chapter in this dissertation.

8.1.1 Chapter 3: Community-Based Event Detection in Temporal Networks

- We show that the proportion of links across communities (during a particular time window) is a detection signature for anomalies over time. We compute the community structure over the aggregate data for a fixed period of time and monitor the number of inter- and intra-community links to estimate the proportion of links across communities. (Section 3.3)

- We illustrate how the diversification of interactions between users (when the information spreads across communities) offers a complementary explanation about why an event becomes an im-

portant topic of global interest that can be used to detect them. To formalize this understanding, we base our reasoning on the principles of diffusion of collective behavior in social networks (i.e., complex contagion) [Weng et al., 2013, Centola and Macy, 2007]. (Section 3.3.4)

- We demonstrate the effectiveness of the proposed method by analyzing the email communication network of Enron [Diesner et al., 2005] (using the events reported in [Diesner et al., 2005, Collingsworth et al., 2009]) and the interactions between Twitter users during the Boston Marathon bombing. (Section 3.4)

8.1.2 Chapter 4: Insider Threat Modeling

- We propose a generic temporal graph analysis framework to model the evolution of bipartite graphs and their equivalent one-mode projection. The proposed framework is based on the idea that the evolution of user-system interactions can be abstracted as a set of consecutive graphs—also called a graph stream. This framework allows the ability to select the granularity of network formation which has been found to be application dependent [Krings et al., 2012]. We use the proposed framework to formalize a set of measurements of the observed graphs at each time interval. (Section 4.3.2)

- We propose a generic framework to compute the performance of an event detector algorithm. (Section 4.3.5)

- We compare the performance of the proposed method with a naive random and edge dependent algorithms. (Section 4.4.4).

- We use graph mining to reveal that some properties of the one-mode projection of the bipartite graph significantly change in the presence of precipitating events. To do this, we leverage more than 22 years of data on user-system interactions in a version control system. In particular, we show that users tend to diversify their patterns of interactions with components after a precipitating event is announced. Our results suggest that this change in user behavior can be used to infer the existence of a threat in a timely manner so that risk mitigation is possible before the insider

completes their actions. Our work is differentiated from the work in [Heymann and Le Grand, 2013] in three ways. First, we rely on the notion of community structure to inform the detection process. Second, we integrate the volume of interactions between users in different communities into the event detection. Finally, we quantify the perturbations inserted in the system after precipitating events that might lead to insider threats. Methodologically closest to our work is an analysis of the Enron email corpus and Twitter data in [Moriano et al., 2017a]. This work is differentiated not only by the domain (i.e., version control system) but also in that we abstract interactions as a bipartite graph and compare our detection results with standard detection approaches. (Section 4.4.2)

- Leveraging the knowledge that insider operations can be motivated by the insiders environment, we illustrate these can be detected by the proposed method. Given the knowledge beyond computer security that the are precipitating events, our results hint at the scale of the insider threats that are not detected using current methods. (Section 4.5)

8.1.3 Chapter 5: Macroeconomics of Routing Anomalies

- We examine routing anomalies as likely due to incompetence, potential ecrime, or intelligence operations using macroeconomics by leveraging three theories from criminology and global measures of technology adoption. The analysis reveals that exports of technology were not statistically significant, undermining the argument for incompetence. However, the results show the correlations that would be expected if these incidents are crime, and motivated by profit. This aligns with hypotheses others have derived from proprietary network data. Specifically, we could not reject the hypotheses that hijacks can be partially explained by guardianship and relative deprivation theories of crime. (Section 5.4.3)

- We show that civil conflict and surveillance are associated with the disproportionate origination of routing anomalies. This supports the possibility of the use of routing anomalies for national intelligence. (Section 5.4.4)

8.1.4 Chapter 6: Characterizing Routing Anomalies Through Graph Mining

- We propose a generic temporal graph analysis framework to model the evolution on the Internet at the AS-level. The proposed framework is based on the idea that the evolution of the Internet can be abstracted as a dynamic system of consecutive graphs—also called graph stream. We use the proposed framework to formalize a set of measurements of the observed graphs at each time instant. (Section 6.3.1.3)

- We use graph mining to reveal that some properties of the AS-level graphs are useful for early detection of routing incidents. In particular, we show that centrality, average path length, and clustering measurements are subject to perturbation when they are analyzed using k-shell decomposition, mainly to decompose graphs between the core and the periphery. Our results suggest that topological signatures from the AS-level graph representation can be used to infer that an anomalous routing event is happening before widespread disruption. (Section 6.3.1.4)

- We study the capabilities of the proposed approach by building AS-level graphs of three different large-scale routing incidents, i.e., Indosat in April 2014, Telecom Malaysia in June 2015, and Bharti Airtel Ltd. in November 2015. (Section 6.3.1.1)

- Our work is differentiated from the work in [Kruegel et al., 2003, Gaertler and Patrignani, 2004] in that we use dynamic update information from the RouteViews project (with granularity every 15 minutes) to reconstruct the network topology at the AS-level and study the robustness of network topological properties, before, during, and after the incident. This approach allows for differentiation between normal behavior at the network level and disruption or anomalous changes during the incidents. The three cases we address were easily identified following the large-scale disruption but not before. (Section 6.4)

8.1.5 Chapter 7: Bursty Announcements for Early Detection of BGP Routing Anomalies

- We validate our conjecture that inter-arrival time patterns of BGP announcements are a useful signature for early identification of large-scale routing incidents. We show that bursty patterns

of announcements are noticeable before the detection of the incidents by the current state-of-the-art detection system [Toonk]. To do so, we quantify the burstiness of BGP announcements by observing that when there are large-scale incidents, there are groups of announcements with short inter-arrival times followed with larger ones. We report that this observation is independent of the volume of announcements. (Section 7.4.2)

- We describe the design of a proof-of-concept BGP anomaly detection method that only uses data from current route collectors. We use RouteViews route collectors to compute a detection signature of large-scale incidents based on the impact of short inter-arrival times. We discuss how it is possible to anticipate more clearly and accurately when an incident is imminent depending on the view of a specific collector (quantified by the number of router feeders). (Section 7.4.3)

- We report results of a longitudinal analysis of large-scale routing incidents. We evaluated the proposed method by studying three different large-scale routing incidents, i.e., Indosat in April 2014, Telecom Malaysia in June 2015, and Bharti Airtel Ltd. in November 2015. Our approach allows for statistically significant differentiation between normal behavior and disruption or anomalous changes during the incidents. Without the method we have developed, when the three cases occurred they were identified only after the large-scale disruption not in the early phases of the attack. (Section 7.3.1.2)

8.2 Future Work

Several projects that are a natural extension of the chapters in this dissertation are presented below. They covered the spectrum from theoretical to applied research. These projects are ordered with respect to their relationship with the content of the chapters presented in this dissertation.

8.2.1 Limits of Community-Based Event Detection

I would like to develop a set of tools and methods to computationally test how the robustness of the community structure of temporal networks can be leveraged for anomaly detection and resilience.

I hypothesize that the community structure of graphs provides a basis for detection of anomalies that is not biased against the density of interactions [Moriano et al., 2017a]. I plan to use benchmark networks with a known community structure, such as the works in [Holland et al., 1983, Lancichinetti et al., 2008], to test the limits of the robustness of the community structure with respect to the random or targeted aggregation of edges. Some works have focused on the robustness of the community structure under random deletion of edges [Karrer et al., 2008, Yan et al., 2018]. However in cybersecurity an unusual volume of interactions may not necessarily represent an anomaly but reflect seasonal behaviors. I expect that this set of computational experiments can measure to what extent the community structure of a system is stable and when its changes are a useful anomaly detection signature.

8.2.2 Understanding Software Quality in Developer-Component Temporal Graphs

I want to characterize the collective characteristics of software developers and their relationship to software failures. There is evidence suggesting that the improvement of software quality reduces the number of security vulnerabilities. I will explore the use of new algorithms that take into account the inner social structure of developer communities and measure organizational context to measure the integrity of their interactions with software components. Specifically, I will apply a range of graph and data analysis techniques to a variety of data (characteristics of individual developers and software components, failures by components, and temporal logs of interactions) to evaluate the efficacy of network analysis in generating probabilistic trust indicators for the resulting code. The proposed analysis will enable an understanding of the necessary conditions for addressing security vulnerabilities (such as those posed by insiders) to improve software quality.

8.2.3 Relationship Between On-Line and Off-Line Cross Country Conflicts

I would like to build statistical models to understand the variation of cyberattacks encountered and hosted by nation-states. This project will include public data about international hostilities [Wilkenfeld et al., 2010] and alliances [Palmer et al., 2015] as this reveals a rich set of interac-

tions between countries in terms of agreements (e.g., treaties, entente agreements, non-aggression treaties, trade, neutrality pacts, and defense spending) that are a basis for examining the relationship between offline and online conflict. In particular, I would like to investigate to what extent offline conflict or cooperation between nations is associated with online conflictive or cooperative international relationships. There are rich data sources about online conflict in different cyberspace realms [Moriano et al., 2017c, Mezzour et al., 2017]. However, the interaction and causalities between these have not been analyzed using temporal graphs of interactions. I expect that changes in the offline interactions between countries may explain phase transition in the dynamics of online interactions. Empirically identifying factors behind such variation can provide a scientific empirical basis to policy actions to reduce cybersecurity incidents among nations and their allies.

8.3 Final Remarks

Real-world relational data is omnipresent in today's world. Interactions derived from such data can be modeled as networks aiming to provide a better understanding of the structure and dynamics of the underlying systems. Understanding the temporal evolution of these networks is crucial to provide a more detailed characterization of the system's function. Irregularities, i.e., anomalies, in the general evolution of these networks are usually associated with undesired and often malicious behavior.

Characterizing regular behavior is often a prerequisite to identify these anomalies. However, variations in the volume of interactions during a system's evolution under particular circumstances may be the norm. Identifying stationary trends that allow us to design reliable detection algorithms remains an open challenge. Identification of anomalous insiders typically relies on supervised learning models that use labeled data. However, such labeled data is not easily obtainable, which opens the door for the use of anomaly detection methods.

By creating hypotheses about the nature of attacks as part of this dissertation, we have developed anomaly detection methods that are robust against the density of interactions in a system—they are not biased against the density of interactions in longitudinal security data. The meth-

ods developed in this dissertation are data and network science centric, including the analysis of the community structure and k-shell decomposition of the underlying graphs, the use of macroeconomic analysis, and the analysis of inter-arrival times of events to detect bursts of malicious route updates. We illustrate the potential of the previously developed methods applied to different real-world scenarios, including email interactions, social media, code repositories, and Internet control-plane updates.

Collecting the appropriate data will continue to be a challenge in these endeavors. Note that for the analysis in Chapter 3, we use a subset of emails in Enron—which comprises communications primarily between managers—and a random sample of tweets using the Twitter streaming API. In Chapter 4, we use logs of a proprietary control version system for a specific codebase. In Chapter 5, we collect data about BGP anomalous incidents relying on the accurate classification of a third-party system and publicly available indicators from the World Bank. In Chapters 6 and 7, we use data collected and served through the BGPStream API. Whether these datasets are enough to identify anomalies in the different security scenarios that we tackle in this dissertation is a matter of further study. This is a difficult challenge because there are policy, privacy, and operational issues in getting the right data.

The findings in this dissertation, however, are not opaque given these constraints. Although data alone do not provide insights to the different problems we study, we have shown that a hypothesis-driven approach—based on mathematical abstractions—with respect to the nature of the attacks informs the process of selecting the more appropriate data. Through that characterization, this dissertation examined the temporal evolution of these system and provided a more nuanced characterization of the interaction of normal functionality and anomalous behavior—usually undesired and often malicious.

Bibliography

V. Chandola, A. Banerjee, and V. Kumar. Anomaly Detection: A Survey. *ACM Computing Surveys (CSUR)*, 41(3):15, 2009.

L. Akoglu, H. Tong, and D. Koutra. Graph based anomaly detection and description: A survey. *Data Mining and Knowledge Discovery*, 29(3):626–688, 2015.

A. Vespignani. Predicting the Behavior of Techno-Social Systems. *Science*, 325(5939):425–428, 2009.

D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. V. Alstyne. Computational Social Science. *Science*, 323(5915):721–723, 2009.

R. Sommer and V. Paxson. Outside the closed world: On using machine learning for network intrusion detection. In *IEEE Symposium on Security and Privacy*, pages 305–316, Oakland, CA, USA, 2010.

T. Ide and H. Kashima. Eigenspace-Based Anomaly Detection in Computer Systems. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 440–449, Seattle, WA, USA, 2004.

J. Neil, C. Hash, A. Brugh, M. Fisk, and C. B. Storlie. Scan Statistics for the Online Detection of Locally Anomalous Subgraphs. *Technometrics*, 55(4):403–414, 2013.

P.-A. Vervier, O. Thonnard, and M. Dacier. Mind Your Blocks: On the Stealthiness of Malicious BGP Hijacks. In *Proceeding of the Network and Distributed System Security Symposium*, San Diego, CA, USA, 2015.

P. Moriano, J. Finke, and Y.-Y Ahn. Community-based anomalous event detection in temporal networks. In *Conference on Complex Systems*, Cancún, Mexico, September 2017a.

P. Moriano, J. Finke, and Y.-Y Ahn. Community-Based Event Detection in Temporal Networks. *Sci. Rep.*, 9(1):4358, 2019a.

P. Moriano, J. Pendleton, S. Rich, and L. J. Camp. Insider Threat Event Detection in User-System Interactions. In *Proceeding of the 9th ACM CCS International Workshop on Managing Insider Security Threats (MIST)*, pages 1–12, Dallas, TX, USA, 2017b.

P. Moriano, J. Pendleton, S. Rich, and L. J. Camp. Stopping the Insider at the Gates: Protecting Organizational Assets Through Graph Mining. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 9(1):4–29, 2018a.

A. Dainotti, C. Squarcella, Aben E, K. C. Claffy, M. Chiesa, M. Russo, and A. Pescapé. Analysis of Country-wide Internet Outages Caused by Censorship. In *Proceedings of the Internet Measurement Conference*, pages 1–18, Berlin, Germany, 2011.

P. Moriano, S. Achar, and L. Jean Camp. Macroeconomic Analysis of Routing Anomalies. In *Conference on Communication, Information and Internet Policy (TPRC)*, Arlington, VA, USA, 2016.

P. Moriano, S. Achar, and L. J. Camp. Incompetents, criminals, or spies: Macroeconomic analysis of routing anomalies. *Comput. Secur.*, 70:319–334, 2017c.

P. Moriano, S. Iyer, and L. J. Camp. Characterization of Internet Routing Anomalies Through Graph Mining. Technical report, Indiana University Technical Report TR733, 2017d.

P. Moriano, R. Hill, and L. J. Camp. Hijacking Network Traffic: Temporal Analysis of Adverse Changes in the Internet Topology. In *Conference on Complex Systems*, Thessaloniki, Greece, September 2018b.

Q. Ding, N. Katenka, P. Barford, E. Kolaczyk, and M. Crovella. Intrusion as (Anti)social Communication: Characterization and Detection. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 886–894, Beijing, China, 2012.

R. J. Bolton and D. J. Hand. Unsupervised Profiling Methods for Fraud Detection. In *Credit Scoring and Credit Control VII*, pages 235–255, Edinburgh, UK, 2001.

L. Invernizzi, P. M. Comparetti, S. Benvenuti, C. Kruegel, M. Cova, and G. Vigna. Evilseed: A guided approach to finding malicious web pages. In *Proceedings of IEEE Symposium on Security and Privacy*, pages 428–442, San Francisco, CA, USA, 2012.

X. Shi, Y. Xiang, Z. Wang, X. Yin, and J. Wu. Detecting Prefix Hijackings in the Internet with Argus. In *Proceedings of the 2012 ACM Conference on Internet Measurement Conference*, pages 15–28, Boston, MA, USA, 2012.

D. Hawkins. *Identification of outliers*, volume 11. Springer, 1980.

T. La Fond, J. Neville, and B. Gallagher. Anomaly Detection in Networks with Changing Trends. In *ACM SIGKDD 2014 ODD Workshop on Outlier Detection & Description under Data Diversity*, pages 3–12, New York, NY, USA, 2014a.

N. Abe, B. Zadrozny, and J. Langford. Outlier Detection by Active Learning. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 504–509, Philadelphia, PA, USA, 2006.

J. H. M. Janssens, I. Flesch, and E. O. Postma. Outlier Detection with One-Class Classifiers from ML and KDD. In *Proceedings of the International Conference on Machine Learning and Applications*, pages 147–153, Miami, FL, USA, 2009.

C. C. Aggarwal and P. S. Yu. Outlier detection for high dimensional data. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 37–46, Santa Barbara, CA, USA, 2001.

X. Wang and X. L. Wang D. M. Wilkes. A Minimum Spanning Tree-Inspired Clustering-Based Outlier Detection Technique. In *Proceedings of the 12th Industrial Conference on Advances in Data Mining: Applications and Theoretical Aspects*, pages 209–223, Berlin, Germany, 2012.

K. Smets and J. Vreeken. The Odd One Out: Identifying and Characterising Anomalies. In *Proceedings of the SIAM International Conference on Data Mining*, pages 804–815, Mesa, AZ, USA, 2011.

M.-L. Shyu, S.-C. Chen, K. Sarinnapakorn, and L. Chang. Principal Component-based Anomaly Detection Scheme. In *Foundations and Novel Approaches in Data Mining*, pages 311–329. Springer, 2006.

M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, 45(2):167–256, 2003.

S. Ranshous, S. Shen, D. Koutra S. Harenberg, C. Faloutsos, and N. F. Samatova. Anomaly detection in dynamic networks: A survey. *Wiley Interdiscip. Rev. Comput. Stat.*, 7(3):223–247, 2015.

J. Sun, C. Faloutsos, S. Papadimitriou, and P. S. Yu. GraphScope: Parameter-Free Mining of Large Time-Evolving Graphs. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 687–696, San Jose, CA, USA, 2007.

J. Rissanen. Modeling by Shortest Data Description. *Automatica*, 14(5):465–471, 1978.

C. C. Noble and D. J. Cook. Graph-Based Anomaly Detection. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 631–636, Washington, DC, USA, 2003.

L. Akoglu and C. Faloutsos. Event Detection in Time Series of Mobile Communication Graphs. In *27th Army Science Conference*, pages 77–79, Orlando, FL, USA, 2010.

R. A. Rossi, B. Gallagher, J. Neville, and K. Henderson. Modeling Dynamic Behavior in Large Evolving Graphs. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, pages 667–676, Rome, Italy, 2013.

K. Henderson, B. Gallagher, T. Eliassi-Rad, H. Tong, S. Basu, L. Akoglu, D. Koutra, C. Faloutsos, and L. Li. RolX: Structural Role Extraction & Mining in Large Graphs. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1231–1239, Beijing, China, 2012.

P. Shoubridge, M. Kraetzl, W. Wallis, and H. Bunke. Detection of Abnormal Change in a Time Series of Graphs. *Journal of Interconnection Networks*, 03(01n02):85–101, 2002.

B. Pincombe. Anomaly Detection in Time Series of Graphs Using ARMA Processes. *ASOR Bulletin*, 24(4), 2005.

D. Koutra, J. Vogelstein, and C. Faloutsos. DeltaCon: A Principled Massive-Graph Similarity Function. In *Proceedings of the 13th SIAM International Conference on Data Mining*, pages 162–170. Austin, TX, USA, 2013.

D. Koutra, T.-Y. Ke, U. Kang, D. H. Chau, H.-K. K. Pao, and C. Faloutsos. Unifying Guilt-by-Association Approaches: Theorems and Fast Algorithms. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 245–260, Athens, Greece, 2011.

C. C. Aggarwal, Y. Zhao, and P. Yu. Outlier Detection in Graph Streams. In *Proceedings of the 27th IEEE International Conference on Data Engineering*, pages 399–409, Hannover, Germany, 2011.

S. Hirose, K. Yamanishi, T. Nakata, and R. Fujimaki. Network Anomaly Detection Based on Eigen Equation Compression. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1185–1194, Paris, France, 2009.

D. Duan, Y. Li, Y. Jin, and Z. Lu. Community Mining on Dynamic Weighted Directed Graphs. In *Proceedings of the 1st ACM International Workshop on Complex Networks Meet Information and Knowledge Management*, pages 11–18, Hong Kong, China, 2009.

L. Peel and A. Clauset. Detecting Change Points in the Large-Scale Structure of Evolving Networks. In *29th AAAI Conference on Artificial Intelligence (AAAI)*, pages 2914–2920. Austin TX, USA, 2015.

T. La Fond, J. Neville, and B. Gallagher. Anomaly detection in dynamic networks of varying size. *arXiv preprint arXiv:1411.3749*, 2014b.

M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.*, 99(12):7821–7826, 2002.

M. E. J. Newman and J. Park. Why social networks are different from other types of networks. *Phys. Rev. E*, 68(3):036122, 2003.

P. Moriano and J. Finke. Characterizing the Relationship Between Degree Distributions and Community Structures. In *Proceedings of the American Control Conference*, pages 2383–2388, Portland, OR, USA, 2014.

J. Kleinberg and S. Lawrence. The Structure of the Web. *Science*, 294(5548):1849–1850, 2001.

E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.*, 10(3):186, 2009.

Q. Song and X. Wang. Efficient Routing on Large Road Networks Using Hierarchical Communities. *IEEE Trans. Intell. Transp. Syst.*, 12(1):132–140, 2011.

F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. Defining and identifying communities in networks. *Proc. Natl. Acad. Sci. U.S.A.*, 101(9):2658–2663, 2004.

M. E. J. Newman. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.*, 103(23):8577–8582, 2006.

V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp*, 2008(10):P10008, 2008.

M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U.S.A.*, 105(4):1118–1123, 2008.

G. Palla, I. Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814, 2005.

Y-Y. Ahn, J. P. Bagrow, and S. Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466(7307):761, 2010.

S. Fortunato. Community detection in graphs. *Phys. Rep.*, 486(3-5):75–174, 2010.

L. Bohlin, D. Edler, A. Lancichinetti, and M. Rosvall. Community Detection and Visualization of Networks with the Map Equation Framework. In *Measuring Scholarly Impact*, pages 3–34. Springer, 2014.

G. Huston. BGP Routing Table Analysis Reports, 2019. URL `http://bgp.potaroo.net/`. Date last accessed November, 30 2018.

K. Lougheed and Y. Rekhter. A Border Gateway Protocol. RFC 1105, RFC Editor, June 1989. Date last accessed August, 28 2017.

Y. Rekhter, T. Li, and S. Hares. A Border Gateway Protocol 4 (BGP-4). RFC 4271, RFC Editor, January 2006. URL `https://tools.ietf.org/html/rfc4271`. Date last accessed August, 28 2017.

V. Fuller and T. Li. Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan. RFC 4632, RFC Editor, August 2006.

J. Mitchell. Autonomous System (AS) Reservation for Private Use. RFC 6996, RFC Editor, July 2013.

Q. Vohra and E. Chen. BGP Support for Four-Octet Autonomous System (AS) Number Space. RFC 6793, RFC Editor, December 2012.

R. Mahajan, D. Wetherall, and T. Anderson. Understanding BGP Misconfiguration. *SIGCOMM Comput. Commun. Rev.*, 32(4):3–16, 2002.

H. Ballani, P. Francis, and Xinyang Zhang. A Study of Prefix Hijacking and Interception in the Internet. *SIGCOMM Comput. Commun. Rev.*, 37(4):265–276, August 2007.

C. Zheng, L. Ji, D. Pei, J. Wang, and P. Francis. A Light-Weight Distributed Scheme for Detecting IP Prefix Hijacks in Real-time. *ACM SIGCOMM Comput. Commun. Rev.*, 37(4):277–288, 2007.

M. Lad, D. Massey, D. Pei, Y. Wu, B. Zhang, and L. Zhang. PHAS: A Prefix Hijack Alert System. In *Proceedings of the 15th Conference on USENIX Security Symposium*, pages 153–166, Vancouver, BC, Canada, 2006.

V. Khare, Q. Ju, and B. Zhang. Concurrent prefix hijacks: Occurrence and impacts. In *Proceedings of the 2012 Internet Measurement Conference*, pages 29–35, Boston, MA, USA, 2012. ISBN 9781450317054.

P. Sermpezis, V. Kotronis, P. Gigis, X. Dimitropoulos, D. Cicalese, A. King, and A. Dainotti. ARTEMIS: Neutralizing BGP Hijacking within a Minute. *IEEE/ACM Trans. Netw.*, 26(6):2471–2486, 2018a.

A. Toonk. BGPmon (commercial). URL `https://bgpmon.net/`. Date last accessed November, 30 2018.

X. Zhao, D. Pei, L. Wang, D. Massey, A. Mankin, S. F. Wu, and L. Zhang. An Analysis of BGP Multiple Origin AS (MOAS) Conflicts. In *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*, pages 31–35, San Francisco, CA, USA, 2001.

P. Sermpezis, V. Kotronis, A. Dainotti, and X. Dimitropoulos. A Survey Among Network Operators on BGP Prefix Hijacking. *ACM SIGCOMM Comput. Commun. Rev.*, 48(1):64–69, 2018b.

Z. Zhang, Y. Zhang, Y. C. Hu, Z. M. Mao, and R. Bush. iSPY: Detecting IP Prefix Hijacking on My Own. *IEEE/ACM Trans. Netw.*, 18(6):1815–1828, 2010.

Y. Xiang, Z. Wang, X. Yin, and J. Wu. Argus: An accurate and agile system to detecting IP prefix hijacking. In *Proceedings of the 19th IEEE International Conference on Network Protocols*, pages 43–48, Vancouver, BC, Canada, 2011.

X. Hu and Z. M. Mao. Accurate Real-Time Identification of IP Prefix Hijacking. In *IEEE Symposium on Security and Privacy*, pages 3–17, Berkeley, CA, USA, 2007.

J. Schlamp, R. Holz, Q. Jacquemart, G. Carle, and E. W. Biersack. HEAP: Reliable Assessment of BGP Hijacking Attacks. *IEEE J. Sel. Areas Commun.*, 34(6):1849–1861, 2016.

S. Kent, C. Lynn, and K. Seo. Secure Border Gateway Protocol (S-BGP). *IEEE J. Sel. Areas Commun.*, 18(4):582–592, 2000.

J. Ng. Extensions to BGP to Support Secure Origin BGP (soBGP). Technical report, RFC Editor, 2004. URL `https://tools.ietf.org/html/draft-ng-sobgp-bgp-extensions-02`. Date last accessed August, 28 2017.

M. Lepinski and K. Sriram. BGPsec Protocol Specification. RFC 18, RFC Editor, February 2017. URL `https://tools.ietf.org/html/draft-lepinski-bgpsec-protocol-00`. Date last accessed August, 28 2017.

M. Lepinski and M. Kent. An Infrastructure to Support Secure Internet Routing. RFC 6480, RFC Editor, February 2012. URL `https://tools.ietf.org/html/rfc6480`. Date last accessed August, 28 2017.

D. Cooper, E. Heilman, K. Brogle, L. Reyzin, and S. Goldberg. On the Risk of Misbehaving RPKI Authorities. In *Proceedings of the 12th ACM Workshop on Hot Topics in Networks*, pages 16:1–16:7, College Park, MD, USA, 2013. ISBN 978-1-4503-2596-7.

M. Wählisch, O. Maennel, and T. C. Schmidt. Towards Detecting BGP Route Hijacking Using the RPKI. *ACM SIGCOMM Comput. Commun. Rev.*, 42(4):103–104, 2012.

P. Gill, M. Schapira, and S. Goldberg. Let the Market Drive Deployment: A Strategy for Transitioning to BGP Security. In *Proceedings of the ACM SIGCOMM 2011 Conference*, pages 14–25, Toronto, Ontario, Canada, 2011.

S. Goldberg. Why Is It Taking So Long to Secure Internet Routing? *Commun. ACM*, 57(10): 56–63, 2014. ISSN 1542-7730.

R. Lychev, S. Goldberg, and M. Schapira. BGP Security in Partial Deployment: Is the Juice Worth the Squeeze? In *ACM SIGCOMM Comput. Commun. Rev.*, volume 43, pages 171–182, 2013.

K. Sriram, D. Montgomery, D. McPherson, E. Osterweil, and B. Dickson. Problem Definition and Classification of BGP Route Leaks. RFC 7908, RFC Editor, June 2016. URL `https://tools.ietf.org/html/rfc7908`. Date last accessed December, 7 2018.

S. Goldberg, M. Schapira, P. Hummon, and J. Rexford. How Secure Are Secure Interdomain Routing Protocols. In *Proceedings of the ACM SIGCOMM 2010 Conference*, pages 87–98, New Delhi, India, 2010.

K. Butler, T. R. Farley, P. McDaniel, and J. Rexford. A Survey of BGP Security Issues and Solutions. *Proc. IEEE*, 98(1):100–122, 2010.

B. Al-Musawi, P. Branch, and G. Armitage. BGP Anomaly Detection Techniques: A Survey. *IEEE Commun. Surv. Tutor.*, 19(1):377–396, 2017.

A. Mitseva, A. Panchenko, and T. Engel. The state of affairs in BGP security: A survey of attacks and defenses. *Comput. Commun.*, 124:45–60, 2018.

C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage. Spamalytics: An Empirical Analysis of Spam Marketing Conversion. In *Proceedings of the 15th ACM Conference on Computer and Communications Security*, pages 3–14, Alexandria, VA, USA, 2008.

N. Christin. Traveling the silk road: A measurement analysis of a large anonymous online marketplace. In *Proceedings of the 22nd international conference on World Wide Web*, pages 213–224, Rio de Janeiro, Brazil, 2013.

Microsoft. Microsoft Security Intelligence Report. `https://www.microsoft.com/en-us/download/details.aspx?id=27605`, 2011.

V. Garg, T. Koster, and L. J. Camp. Cross-country analysis of spambots. *EURASIP Journal of Information Security*, 3:1–13, 2013.

M. van Eeten and J. M. Bauer. Economics of Malware: Security Decisions, Incentives and Externalities. In *OECD Science, Technology and Industry Working Papers, No. 2008/01*, 2008.

V. Garg and L. J. Camp. Macroeconomic Analysis of Malware. In *Network and Distributed System Security Symposium Extended Abstracts*, San Diego, CA, USA, 2013.

S. Afroz, V. Garg, D. McCoy, and R. Greenstadt. Honor Among thieves: A Common's Analysis of Cybercrime Economies. In *eCrime Researchers Summit*, pages 1–11, San Francisco, CA, USA, 2013.

T. Moore, R. Clayton, and R. Anderson. The Economics of Online Crime. *J. Econ. Perspect.*, 23 (3):3–20, 2009.

M. van Eeten, J. M. Bauer, H. Asghari, S. Tabatabaie, and D. Rand. The Role of Internet Service Providers in Botnet Mitigation: An Empirical Analysis Based on Spam Data. In *The Ninth Workshop on the Economics of Information Securitty*, Cambridge, MA, USA, 2010.

R. Clayton. Badness in the RIPE Database. `https://www.lightbluetouchpaper.org/2015/10/02/badness-in-the-ripe-database/`, October 2015.

A. Ramachandran and N. Feamster. Understanding the Network-level Behavior of Spammers. *SIGCOMM Comput. Commun. Rev.*, 36(4):291–302, 2006.

H. Asghari, M. J. G. van Eeten, and M. L. Mueller. Unravelling the economic and political drivers of deep packet inspection. An empirical study of DPI use by broadband operators in 75 countries. In *7th Annual Global Internet Governance Academic Network Symposium*, (GigaNet), Baku, Azerbaijan, 2012.

R. Hiran, N. Carlsson, and P. Gill. Characterizing Large-scale Routing Anomalies: A Case Study of the China Telecom Incident. In *Proceedings of the 14th International Conference on Passive and Active Measurement*, pages 229–238, Hong Kong, China, 2013.

A. Toonk. Chinese ISP hijacks the Internet, April 2010. URL `http://www.bgpmon.net/chinese-isp-hijacked-10-of-the-internet/`.

A. Arnbak and S. Goldberg. Loopholes for Circumventing the Constitution: Unrestrained Bulk Surveillance on Americans by Collecting Network Traffic Abroad. *Mich. Telecomm. & Tech. L. Rev.*, 21(2):317–361, 2015.

Kevin Benton and L. Jean Camp. Preventing data exfiltration via political and geographic routing policies. *Available at SSRN 2753133*, 2016.

Dyn Guest Blogs. The New Threat: Targeted Internet Traffic Misdirection, November 2013. URL `https://dyn.com/blog/mitm-internet-hijacking/`. Date last accessed August, 28 2017.

D. M. Cappelli, A. P. Moore, and R. F. Trzeciak. *The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes (Theft, Sabotage, Fraud)*. Addison-Wesley Professional, 1st edition, 2012.

M. Bishop, H. M. Conboy, H. Phan, B. I. Simidchieva, G. S. Avrunin, L. A. Clarke, L. J. Osterweil, and S. Peisert. Insider Threat Identification by Process Analysis. In *IEEE Security and Privacy Workshops*, pages 251–264, San Jose, CA, USA, 2014.

J. R. C. Nurse, O. Buckley, P. A. Legg, M. Goldsmith, S. Creese, G. R. T. Wright, and M. Whitty. Understanding Insider Threat: A Framework for Characterising Attacks. In *IEEE Security and Privacy Workshops*, pages 214–228, San Jose, CA, USA, 2014.

D. Liu, X. Wang, and L. J. Camp. Game-theoretic modeling and analysis of insider threats. *Int. J. Crit. Infr. Prot.*, 1:75–80, 2008.

D. Liu, X. Wang, and L. J. Camp. Mitigating inadvertent insider threats with incentives. In *International Conference on Financial Cryptography and Data Security*, pages 1–16, Accra Beach, Barbados, 2009.

W. Eberle, J. Graves, and L. Holder. Insider Threat Detection Using a Graph-Based Approach. *J. Appl. Secur. Res.*, 6(1):32–81, 2010.

W. Eberle and L. Holder. Scalable anomaly detection in graphs. *Intell. Data Anal/*, 19(1):57–74, 2015.

A. D. Kent, L. M. Liebrock, and J. C. Neil. Authentication graphs: Analyzing user behavior within an enterprise network. *Comput. Secur.*, 48:150–166, 2015.

Y. Chen, S. Nyemba, W. Zhang, and B. Malin. Specializing network analysis to detect anomalous insider actions. *Security Informatics*, 1(1):5, 2012.

W.-K. Wong, A. W. Moore, G. F. Cooper, and M. M Wagner. Bayesian Network Anomaly Pattern Detection for Disease Outbreaks. In *Proceedings of the 20th International Conference on Machine Learning*, pages 808–815, Washington, DC, USA, 2003.

S. Basu and M. Meckesheimer. Automatic outlier detection for time series: an application to sensor data. *Knowl. Inform. Syst.*, 11(2):137–154, 2007.

S. Lin and D. Brown. Criminal Incident Data Association Using the OLAP Technology. *International Conference on Intelligence and Security Informatics*, pages 960–960, 2003.

B. Karrer, E. Levina, and M. E. J. Newman. Robustness of community structure in networks. *Phys. Rev. E*, 77:046119, 2008.

Y. Yang, Z. Li, Y. Chen, X. Zhang, and S. Wang. Improving the Robustness of Complex Networks with Preserving Community Structure. *PLoS One*, 10(2):1–14, 2015.

L. Weng, F. Menczer, and Y.-Y. Ahn. Virality Prediction and Community Structure in Social Networks. *Sci. Rep.*, 3(1):2522, 2013.

J. Diesner, T. L. Frantz, and K. M. Carley. Communication Networks from the Enron Email Corpus "It's Always About the People. Enron is no Different". *Comput. Math. Organ. Theory*, 11(3): 201–228, 2005.

G. Wilson and W. Banzhaf. Discovery of Email Communication Networks from the Enron Corpus with a Genetic Algorithm using Social Network Analysis. In *IEEE Congress on Evolutionary Computation*, pages 3256–3263, Trondheim, Norway, 2009.

J. Sutton, E. S. Spiro, S. Fitzhugh, B. Johnson, B. Gibson, and C. T. Butts. Terse Message Amplification in the Boston Bombing Response. In *Proceedings of the 11th International ISCRAM Conference*, pages 612–621. University Park, PA, USA, 2014.

K. Starbird, J. Maddock, M. Orand, P. Achterman, and R. M. Mason. Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing. In *iConference 2014 Proceedings*, pages 654–662, Berlin, Germany, 2014.

M. E. J. Newman. Detecting community structure in networks. *Eur. Phys. J. B*, 38(2):321–330, 2004.

R. Marks. Enron Timeline. `http://www.agsm.edu.au/bobm/teaching/BE/Enron/timeline.html`, April 2010. Last Accessed: April 10, 2017.

R. K. Darst, C. Granell, A. Arenas, S. Gómez, J. Saramäki, and S. Fortunato. Detection of timescales in evolving complex systems. *Sci. Rep.*, 6:39713, 2016.

M. W. Berry and M. Browne. The 2001 annotated (by topic) Enron email data set. `http://www.cis.jhu.edu/~parky/Enron/Anno_Topic_exp_LDC.pdf`, April 2010. Last Accessed: April 11, 2017.

T. Fawcett. An introduction to ROC analysis. *Pattern Recognit. Lett*, 27(8):861–874, 2006.

M. Gordon and M. Kochen. Recall-precision trade-off: A derivation. *J. Assoc. Inf. Sci. Technol*, 40(3):145–151, 1989.

T. Saito and M. Rehmsmeier. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS One*, 10(3):1–21, 03 2015.

D. Jackson. AP Twitter feed hacked; no attack at White House. `https://www.usatoday.com/story/theoval/2013/04/23/obama-carney-associated-press-hack-white-house/2106757/`, April 2013. Last Accessed: September 25, 2017.

Reuters. Ex-Ford engineer sentenced for trade secrets theft. `http://www.reuters.com/article/us-djc-ford-tradesecrets-idUSTRE73C3FG20110413`, April 2011. Date last accessed July 5, 2017.

FBI. Fannie Mae Corporate Intruder Sentenced to Over Three Years in Prison for Attempting to Wipe Out Fannie Mae Financial Data . https://archives.fbi.gov/archives/baltimore/press-releases/2010/ba121710.htm, December 2010. Date last accessed July 5, 2017.

J. Edwards and M. Hoosenball. NSA contractor charged with stealing secret data. http://www.reuters.com/article/us-usa-cybersecurity-arrest-idUSKCN12520Y, October 2016. Date last accessed July 5, 2017.

D. Culp. Lessons not learned: Insider threats in pathogen research. http://thebulletin.org/lessons-not-learned-insider-threats-pathogen-research, April 2013. Date last accessed July 5, 2017.

Ponemon Institute. 2016 Cost of Cyber Crime Study & the Risk of Business Innovation. Technical report, http://www.ponemon.org/library/2016-cost-of-cyber-crime-study-the-risk-of-business-innovation, 2016. Date last accessed July 5, 2017.

M. E. J. Newman. *Networks: An introduction*. Oxford University Press, 1st edition, 2010.

P. Holme and J. Saramäki. Temporal networks. *Phys. Rep.*, 519(3):97–125, 2012.

P. Parveen, J. Evans, B. Thuraisingham, K. W. Hamlen, and L. Khan. Insider Threat Detection Using Stream Mining and Graph Mining. In *IEEE Third International Conference on Privacy, Security, Risk and Trust and IEEE Third Inernational Conference on Social Computing*, pages 1102–1110, Boston, MA, USA, 2011.

A. P. Moore, D. M. Cappelli, and R. F. Trzeciak. *The "Big Picture" of Insider IT Sabotage Across U.S. Critical Infrastructures*, pages 17–52. Springer US, Boston, MA, USA, 2008.

I. Barnett and J.-P. Onnela. Change point detection in correlation networks. *Sci. Rep.*, 6:18893, 2016.

S. Heymann and B. Le Grand. Monitoring user-system interactions through graph-based intrinsic dynamics analysis. In *IEEE Seventh International Conference on Research Challenges in Information Science*, pages 1–10, Paris, France, 2013.

T. Zhou, J. Ren, M. Medo, and Y.-C. Zhang. Bipartite network projection and personal recommendation. *Phys. Rev. E*, 76(4):046115, 2007.

T. Rashid, I. Agrafiotis, and J. R. C. Nurse. A new take on detecting insider threats: Exploring the use of hidden markov models. In *Proceedings of the 8th ACM CCS International Workshop on Managing Insider Security Threats*, pages 47–56, 2016.

S. Fortunato and D. Hric. Community detection in networks: A user guide. *Phys. Rep.*, 659:1–44, 2016.

J.-L. Guillaume and M. Latapy. Bipartite Structure of All Complex Networks. *Inform. Process. Lett.*, 90(5):215–221, 2004.

iDatalabs. Companies using IBM Rational ClearCase. `https://idatalabs.com/tech/products/ibm-rational-clearcase`, June 23 2017. Date last accessed June 28, 2017.

L. Benamara and C. Magnien. Estimating properties in dynamic systems: The case of churn in p2p networks. In *INFOCOM IEEE Conference on Computer Communications Workshops*, pages 1–6, San Diego, CA, USA, 2010.

J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, pages 177–187, Chicago, Illinois, USA, 2005.

S. Mishra and G. Dhillon. Information Systems Security Governance Research: A Behavioral Perspective. In *1st Annual Symposium on Information Assurance, Academic Track of 9th Annual NYS Cyber Security Conference*, pages 27–35, New York, NY, USA, 2006.

M. Warkentin and R. Willison. Behavioral and policy issues in information systems security: the insider threat. *Eur. J. Inform. Syst.*, 18(2):101–105, 2009.

F. L. Greitzer, P. Paulson, L. Kangas, L. R. Franklin, T. W. Edgar, and D. A. Frincke. Predictive Modelling for Insider Threat Mitigation. *Pacific Northwest National Laboratory, Richland, WA, Tech. Rep. PNNL Technical Report PNNL-65204*, 2009.

Andrew P Moore, David A Mundie, and Matthew L Collins. A System Dynamics Model for Investigating Early Detection of Insider Threat Risk. In *Conference Proceedings of the 31st International Conference of the System Dynamics Society*, pages 978–1, Cambridge, MA, USA, 2013.

T. Benjaminsen. The Norwegian Downsizing Approach in Terms of the Insider Threat-An interpretive study. Master's thesis, Norwegian University of Science and Technology, 2017.

Z. Dong, V. Garg, L. J. Camp, and A. Kapadia. Pools, clubs and security: designing for a party not a person. In *Proceedings of the 2012 New Security Paradigms Workshop*, pages 77–86, Bertinoro, Italy, 2012.

T. C. Pratt and F. T. Cullen. Assessing Macro-Level Predictors and Theories of Crime: A Meta-Analysis. *Crime and Justice*, 32:373–450, 2005.

L. A. Ika, A. Diallo, and D. Thuillier. Critical success factors for World Bank projects: An empirical investigation. *Int. J. Proj. Manag.*, 30(1):105–116, 2012.

M. Felson and L. E. Cohen. Human ecology and crime: A routine activity approach. *Hum. Ecol.*, 8(4):389–406, 1980.

J. R. Blau and P. M. Blau. The cost of inequality: Metropolitan structure and violent crime. *Am. Sociol. Rev.*, 47(1):114–129, 1982.

F. T. Cullen. Social support as an organizing concept for criminology: Presidential address to the academy of criminal justice sciences. *Justice Quarterly*, 11(4):527–559, 1994.

D. Kaufmann, A. Kraay, and M. Mastruzzi. The Worldwide Governance Indicators: Methodology and Analytical Issues. *Hague Journal on the Rule of Law*, 3(2):220–246, 2011.

A. Toonk. Massive route leak causes Internet slowdown, June 2015a. URL `https://bgpmon.net/massive-route-leak-cause-internet-slowdown/`. Date last accessed August, 28 2017.

A. Toonk. How the Internet in Australia went down under, February 2012. URL `http://www.bgpmon.net/how-the-internet-in-australia-went-down-under/`.

Anti-Phishing Working Group (APWG). Phishing Activity Trends Report. `https://docs.apwg.org/reports/apwg_trends_report_q4_2014.pdf`, April 2015.

Cisco Systems. Annual Security Report. `http://www.cisco.com/web/offer/gist_ty2_asset/Cisco_2015_ASR.pdf`, 2015.

L. Quan, J. Heidemann, and Y. Pradkin. When the Internet Sleeps: Correlating Diurnal Networks with External Factors. In *Proceedings of the 2014 ACM Conference on Internet Measurement Conference*, pages 87–100, Vancouver, BC, Canada, 2014.

CAIDA. The CAIDA UCSD AS to Organization Mapping Dataset. `http://www.caida.org/data/as_organizations.xml`, July 2015.

D. Madory. The Vast World of Fraudulent Routing. `http://research.dyn.com/2015/01/vast-world-of-fraudulent-routing/`, January 2015a.

A. F. Zuur, E. N. Ieno, and C. S. Elphick. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.*, 1(1):3–14, 2010.

A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

F. J. Massey. The Kolmogorov-Smirnov Test for Goodness of Fit. *J. Am. Stat. Assoc.*, 46(253): 68–78, 1951.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, 2 edition, 2016.

S. Gallagher. Pirates hack into shipping company's servers to identify booty. `http://arstechnica.com/security/2016/03/pirates-hack-into-shipping-companys-servers-to-identify-booty/`, March 2016.

M. Mitzenmacher. A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Mathematics*, 1(2):226–251, 2004.

B. Edwards, S. Hofmeyr, and S. Forrest. Hype and Heavy Tails: A Closer Look at Data Breaches. In *The 14th Workshop on the Economics of Information Security*, Delft, Netherlands, 2015.

J. M. Samuels. Size and the Growth of Firms. *Rev. Econ. Stud.*, 32(2):105–112, 1965.

G.-.Q. Zhang, G.-Q. Zhang, Q.-F. Yang, S-Q. Cheng, and T. Zhou. Evolution of the Internet and its cores. *New J. Phys.*, 10(12):123027, 2008.

B. Edwards, S. Hofmeyr, G. Stelle, and S. Forrest. Internet Topology over Time. arXiv preprint arXiv:1202.3993v1, 2012.

M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *ACM SIGCOMM Comput. Commun. Rev.*, volume 29, pages 251–262, 1999.

R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378, 2000.

S. Zhou and R. J. Mondragón. The Rich-Club Phenomenon in the Internet Topology. *IEEE Commun. Lett.*, 8(3):180–182, 2004.

C. Kruegel, D. Mutz, W. Robertson, and F. Valeur. *Recent Advances in Intrusion Detection*, volume 2820, chapter Topology-Based Detection of Anomalous BGP Messages, pages 17–35. Springer Berlin Heidelberg, 2003.

M. Gaertler and M. Patrignani. Dynamic analysis of the autonomous system graph. In *International Workshop on Inter-domain Performance and Simulation*, pages 13–24, Budapest, Hungary, 2004.

R. Pastor-Satorras and A. Vespignani. *Evolution and structure of the Internet: A statistical physics approach*. Cambridge University Press, 2007.

E. Zmijewski. Indonesia Hijacks the World, April 2014. URL `https://dyn.com/blog/indonesia-hijacks-world/`. Date last accessed August, 28 2017.

A. Toonk. Large scale BGP hijack out of India, November 2015b. URL `https://bgpmon.net/large-scale-bgp-hijack-out-of-india/`. Date last accessed August, 28 2017.

C. Orsini, A. King, D. Giordano, V. Giotsas, and A. Dainotti. BGPStream: A Software Framework for Live and Historical BGP Data Analysis. In *Proceedings of the 2016 Internet Measurement Conference*, pages 429–444, Santa Monica, CA, USA, 2016.

R. Mazloum, M.-O. Buob, J. Auge, B. Baynat, D. Rossi, and T. Friedman. Violation of Interdomain Routing Assumptions. In *Proceedings of the 15th International Conference on Passive and Active Measurement*, pages 173–182, Los Angeles, CA, USA, 2014.

D. Meyer. University of Oregon Route Views Archive Project, June 2004. URL `http://archive.routeviews.org`. Date last accessed August, 28 2017.

RIPE NCC. RIS Raw Data, February 3 2011. URL `https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/ris-raw-data`. Date last accessed August, 28 2017.

E. Gregori, A. Improta, L. Lenzini, L. Rossi, and L. Sani. On the Incompleteness of the AS-level Graph: A Novel Methodology for BGP Route Collector Placement. In *Proceedings of the 2012 ACM Conference on Internet Measurement Conference*, pages 253–264, Boston, Massachusetts, USA, 2012. ISBN 978-1-4503-1705-4.

K. Chen, C. Hu, W. Zhang, Y. Chen, and B. Liu. On the Eyeshots of BGP Vantage Points. In *Proceedings of the IEEE Conference on Global Telecommunications*, pages 3558–3563, Honolulu, Hawaii, USA, 2009. ISBN 978-1-4244-4147-1.

H. Haddadi, D. Fay, A. Jamakovic, O. Maennel, A. W. Moore, R. Mortier, M. Rio, and S. Uhlig. Beyond Node Degree: Evaluating AS Topology Models. arXiv preprint arXiv:0807.2023v1, 2008.

S. Hofmeyr, T. Moore, S. Forrest, B. Edwards, and G. Stelle. Modeling Internet-Scale Policies for Cleaning up Malware. In *Economics of Information Security and Privacy III*, pages 149–170. Springer, 2013.

J. C. Doyle, D. L. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger. The "robust yet fragile" nature of the Internet. *Proc. Natl. Acad. Sci. U.S.A.*, 102(41):14497–14502, 2005.

J. Yang and J. Leskovec. Overlapping Communities Explain Core-Periphery Organization of Networks. *Proc. IEEE*, 102(12):1892–1902, 2014.

J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani. K-core decomposition of internet graphs: hierarchies, self-similarity and measurement biases. *Netw. Heterog. Media*, 3(2): 371–393, 2008.

N. Lahav, B. Ksherim, E. Ben-Simon, A. Maron-Katz, R. Cohen, and S. Havlin. K-shell decomposition reveals hierarchical cortical organization of the human brain. *New J. Phys.*, 18(8):083013, 2016.

B. Pittel, J. Spencer, and N. Wormald. Sudden Emergence of a Giant k-Core in a Random Graph. *J. Combin. Theory Ser. B*, 67(1):111–151, 1996.

J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani. Large scale networks fingerprinting and visualization using the k-core decomposition. In *Advances in Neural Information Processing Systems 18*, pages 41–50, 2005.

M. Lad, R. Oliveira, B. Zhang, and L. Zhang. Understanding Resiliency of Internet Topology against Prefix Hijack Attacks. In *Proceedings of the 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, pages 368–377, Edinburgh, UK, June 2007.

R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. The (in)Completeness of the Observed Internet AS-level Structure. *IEEE/ACM Trans. Netw.*, 18(1):109–122, 2010.

L. Wang, X. Zhao, D. Pei, R. Bush, D. Massey, A. Mankin, S. F. Wu, and L. Zhang. Observation and Analysis of BGP Behavior Under Stress. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurment*, pages 183–195, Marseille, France, 2002.

J. Li, M. Guidero, Z. Wu, E. Purpus, and T. Ehrenkranz. BGP Routing Dynamics Revisited. *ACM SIGCOMM Comput. Commun. Rev.*, 37(2):5–16, 2007.

P. Moriano, R. Hill, and L. J. Camp. Using Bursty Announcements for Early Detection of BGP Routing Anomalies. Technical report, Under review in SIGCOMM, 2019b.

Dyn Guest Blogs. Pakistan hijacks YouTube. `https://dyn.com/blog/pakistan-hijacks-youtube-1/`, February 2008. Date last accessed August, 28 2017.

Y. Sun, A. Edmundson, L. Vanbever, O. Li, J. Rexford, M. Chiang, and P. Mittal. RAPTOR: Routing Attacks on Privacy in Tor. In *Proceedings of the 25th Conference on USENIX Security Symposium*, pages 271–286, Washington, DC, USA, 2015.

A. Shaw. Spam? Not Spam? Tracking a hijacked Spamhaus IP. `https://greenhost.nl/2013/03/21/spam-not-spam-tracking-hijacked-spamhaus-ip/`, March 2013. Date last accessed August, 28 2017.

A. Peterson. Researchers say U.S. Internet traffic was re-routed through Belarus. That's a problem., November 2013. URL `https://www.washingtonpost.com/news/the-switch/wp/2013/11/20/`

researchers-say-u-s-internet-traffic-was-re-routed-through-belarus-thats-
Date last accessed September 26, 2016.

A. Toonk. Hijack event today by Indosat, April 2014. URL http://bgpmon.net/hijack-event-today-by-indosat/. Date last accessed August, 28 2017.

C. Hall, D. Yu, Z. Zhang, J. Stout, A. Odlyzko, A. W. Moore, L. J. Camp, K. Benton, and R. Anderson. Collaborating with the enemy on network management. In *Cambridge International Workshop on Security Protocols*, pages 154–162, Cambridge, United Kingdom, 2014.

D. Madory. Global Collateral Damage of TMnet leak, June 2015b. URL https://dyn.com/blog/global-collateral-damage-of-tmnet-leak/. Date last accessed August, 28 2017.

H. S. Javitz and A. Valdes. The NIDES statistical component: Description and justification. Technical report, Computer Science Laboratory, SRI International, Menlo Park, CA, March 1993.

H. Yan, R. Oliveira, K. Burnett, D. Matthews, L. Zhang, and D. Massey. BGPmon: A Real-Time, Scalable, Extensible Monitoring System. In *Cybersecurity Applications Technology Conference for Homeland Security*, pages 212–223, Washington, DC, USA, 2009.

Packet Clearing House. Packet Clearing House. URL https://www.pch.net/. Date last accessed November, 30 2018.

E. Gregori, A. Improta, L. Lenzini, L. Rossi, and L. Sani. A Novel Methodology to Address the Internet AS-Level Data Incompleteness. *IEEE/ACM Trans. Netw.*, 23(4):1314–1327, 2015.

S. Murphy. Routing Security and RPKI, November 2015. URL https://www.nanog.org/sites/default/files/04-Murphy-StLouis.pdf. Date last accessed August, 28 2017.

R. Harang and A. Kott. Burstiness of Intrusion Detection Process: Empirical Evidence and a Modeling Approach. *IEEE Trans. Inf. Forensic Secur.*, 12(10):2348–2359, 2017.

K.-I. Goh and A.-L. Barabási. Burstiness and memory in complex systems. *EPL*, 81(4):48002, 2008.

M. Lad, X. Zhao, B. Zhang, D. Massey, and L. Zhang. Analysis of BGP Update Surge During Slammer Worm Attack. In *International Workshop on Distributed Computing*, pages 66–79, Kolkata, India, 2003. Springer.

S. Deshpande, M. Thottan, and B. Sikdar. Early Detection of BGP Instabilities Resulting from Internet Worm Attacks. In *Proceedings of the IEEE Global Telecommunications Conference*, volume 4, pages 2266–2270, Dallas, TX, USA, 2004.

K. Zhang, A. Yen, X. Zhao, D. Massey, S. F. Wu, and L. Zhang. On Detection of Anomalous Routing Dynamics in BGP. In *International Conference on Research in Networking*, pages 259–270, Athens, Greece, 2004. Springer.

E. Gregori, A. Improta, and L. Sani. On the African peering connectivity revealable via BGP route collectors. In *International Conference on e-Infrastructure and e-Services for Developing Countries*, pages 368–376, Lagos, Nigeria, 2017. Springer.

A.-L. Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207, 2005.

E.-K. Kim and H.-H. Jo. Measuring burstiness for finite event sequences. *Phys. Rev. E*, 94:032311, 2016.

CAIDA AS Rank, Juy 2018. URL `http://as-rank.caida.org/`. Date last accessed November, 30 2018.

C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian. Delayed Internet Routing Convergence. *ACM SIGCOMM Comput. Commun. Rev.*, 30(4):175–187, 2000.

M. Apostolaki, A. Zohar, and L. Vanbever. Hijacking Bitcoin: Routing Attacks on Cryptocurrencies. In *IEEE Symposium on Security and Privacy*, pages 375–392, San Jose, CA, USA, 2017.

A. Lutu, M. Bagnulo, and R. Stanojevic. An economic side-effect for prefix deaggregation. In *Proceedings IEEE INFOCOM Workshops*, pages 190–195, Orlando, FL, USA, 2012.

D. Centola and M. Macy. Complex Contagions and the Weakness of Long Ties. *Am. J. Sociol.*, 113(3):702–734, 2007.

B. Collingsworth, R. Menezes, and P. Martins. Assessing organizational stability via network analysis. In *IEEE Symposium on Computational Intelligence for Financial Engineering (CIFEr)*, pages 43–50, Nashville, TN, USA, 2009.

G. Krings, M. Karsai, S. Bernhardsson, V. D. Blondel, and J. Saramäki. Effects of time window size and placement on the structure of an aggregated communication network. *EPJ Data Science*, 1 (1):4, May 2012.

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Soc. Networks*, 5(2):109 – 137, 1983.

A. Lancichinetti, S. Fortunato, and F. Radicchi. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E*, 78:046110, 2008.

X. Yan, L. G. S. Jeub, A. Flammini, F. Radicchi, and S. Fortunato. Weight thresholding on complex networks. *Phys. Rev. E*, 98:042304, 2018.

J. Wilkenfeld, M. Brecher, J. Hewitt, K. Beardsley, and P. Eralp. International Crisis Behavior Datasets. http://www.icb.umd.edu/dataviewer/, July 2010. Date last accessed October, 16 2018.

G. Palmer, V. d'Orazio, M. Kenwick, and M. Lane. The MID4 dataset, 2002–2010: Procedures, coding rules and description. *Conflict Manag. Peace*, 32(2):222–242, 2015.

G. Mezzour, K. M. Carley, and L. R. Carley. Global Variation in Attack Encounters and Hosting. In *Proceedings of the Hot Topics in Science of Security: Symposium and Bootcamp*, HoTSoS, pages 62–73, New York, NY, USA, 2017. ACM.

A Additional Topological Properties BGP Graph Mining

In this section, we provide more details on the additional empirical measurements we computed to test the proposed hypothesis.

A.1 Global Structure

A.1.1 An Indonesian ISP Hijacking the World

Here, we report on the results of centrality measures which illustrate the prominence of ASes. Figure A.1 shows the number of nodes over time. From this plot, it is possible to infer that the only significant change in this measure is for the graph that is captured at April 2, 2014, at 12:00 and April 3, 2014, at 6:00.

To better understand this behavior, we also study the dynamic transition of the number of edges in Figure A.2. We observe that there is a considerable decrease in the number of edges for the graphs that are built on the same snapshots—in accordance with the measure of the number of nodes.

Figure A.3 shows the number of nodes at various k-levels of cores graphs, i.e., $k = 1, 10$, and the maximum k possible—the one that encloses the nucleus of the Internet. As we might expect, it is not possible to observe significant changes regarding the total number of nodes in this time series. This suggests that the core remains almost the same with respect to the number of ASes.

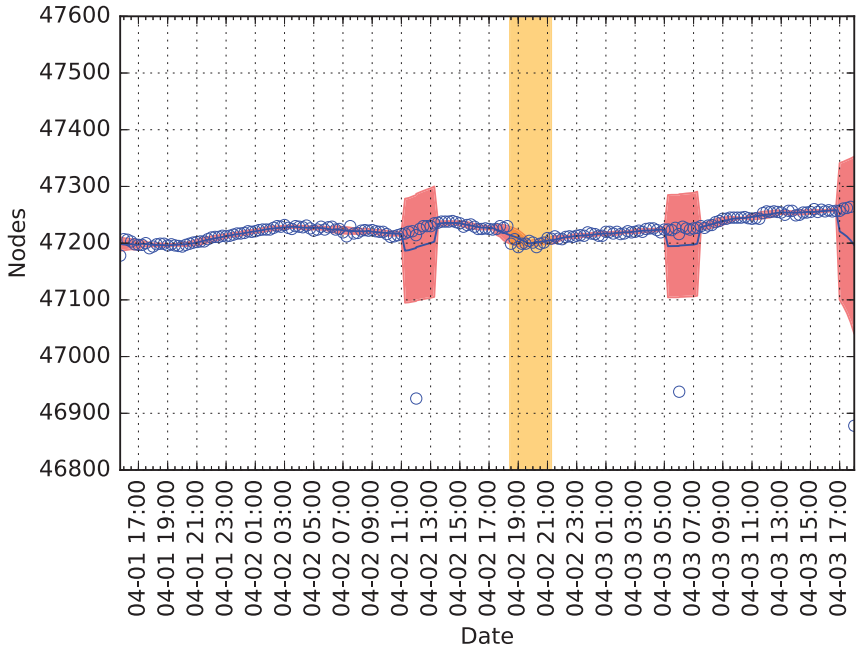

Figure A.1: Number of nodes Indonesia event.

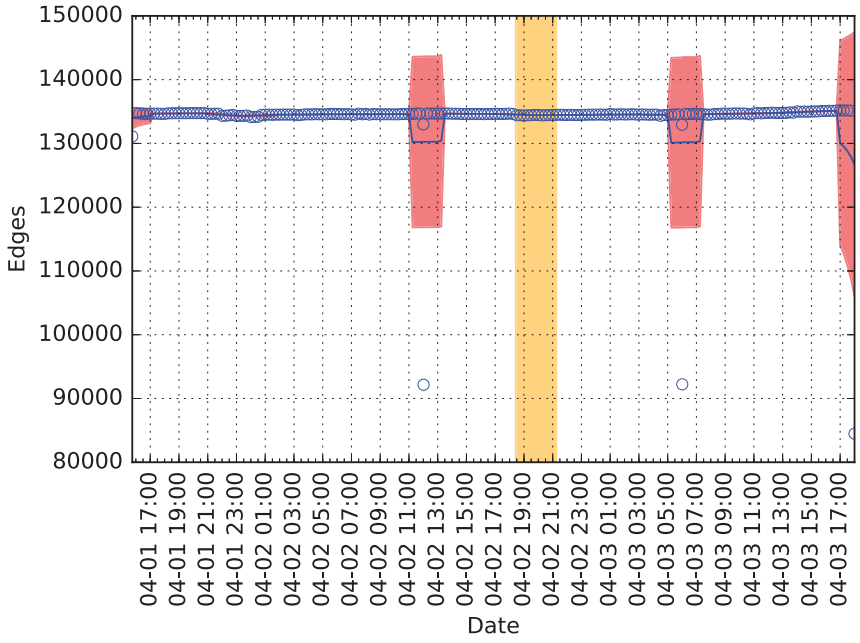

Figure A.2: Number of edges Indonesia event.

A.1.2 Global Collateral Damage of Telecom Malaysia Leak

For this incident, Figures A.4, A.5 illustrate general centrality measures for the number of nodes and edges, respectively. As can be seen, the only significant variations for these properties occur

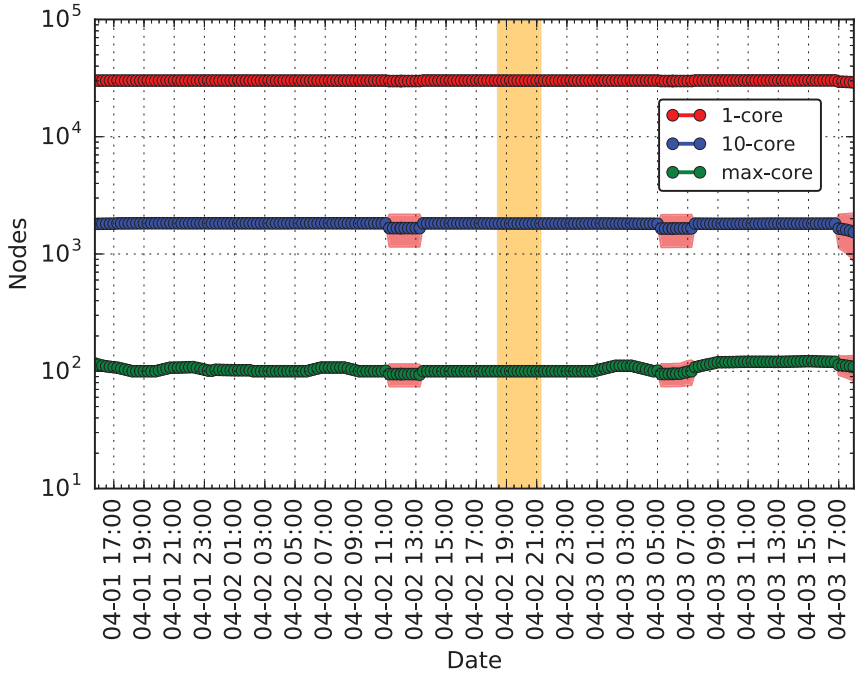

Figure A.3: Nodes per core Indonesia event.

for observations derived at June 11, 2015, at 10:00, and 18:00; June 15, 2014, at 00:00, and 18:00; and June 13, 2015, at 00:00, and 8:00. Similarly, for the core graphs, we computed the number of nodes as is shown in Figure A.6. We did not observe significant variations for these properties.

A.1.3 Large Scale BGP Hijack in India

For the Indian incident, we tracked of the same properties we did for the Indonesian and Malaysian incidents. In particular, Figures A.7, A.8 show the number of nodes and edges, for each graph snapshot during the observation period. As in the previous cases, there are some discontinuities in each of these time series. The discontinuities are evident in November 5, 2015, at 10:00, 12:00, 14:00, 16:00, 18:00, 20:00, 22:00; and November 6, 2015, at 00:00, 02:00, 04:00, and 06:00. Figure A.9 shows the number of nodes in the core subgraphs. This measure does not reveal significant changes during the observation period.

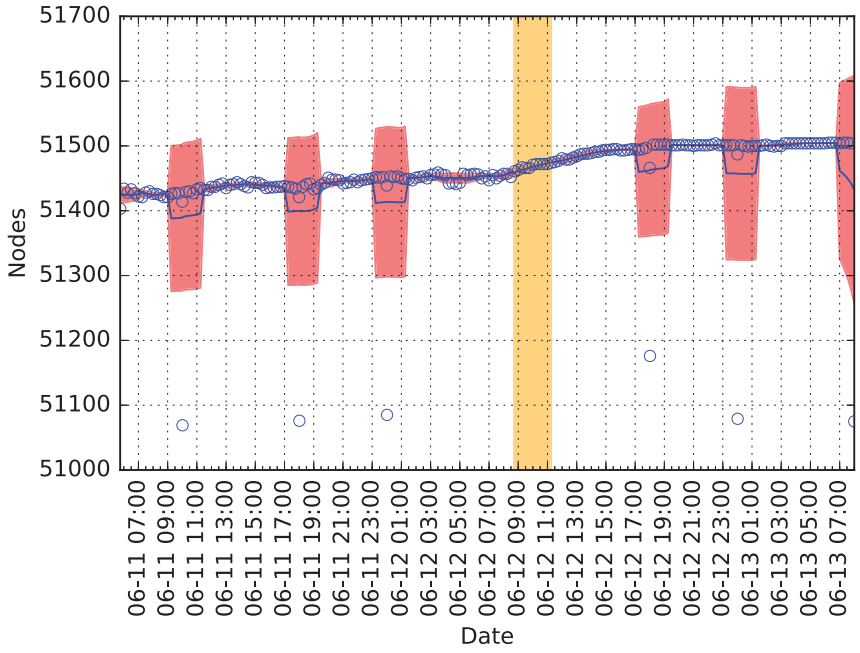

Figure A.4: Number of nodes Malaysia event.

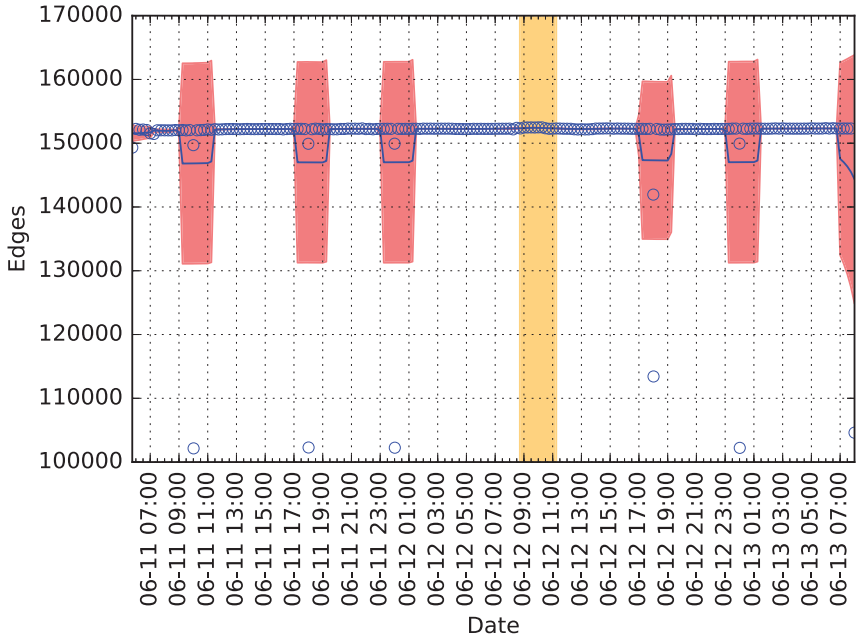

Figure A.5: Number of edges Malaysia event.

A.2 Community Structure

A.2.1 An Indonesian ISP Hijacking the World

Figure A.10 shows the average clustering coefficient of the graph snapshots during the period of study. This measure seems to be stable during the observation period except for the discontinuities

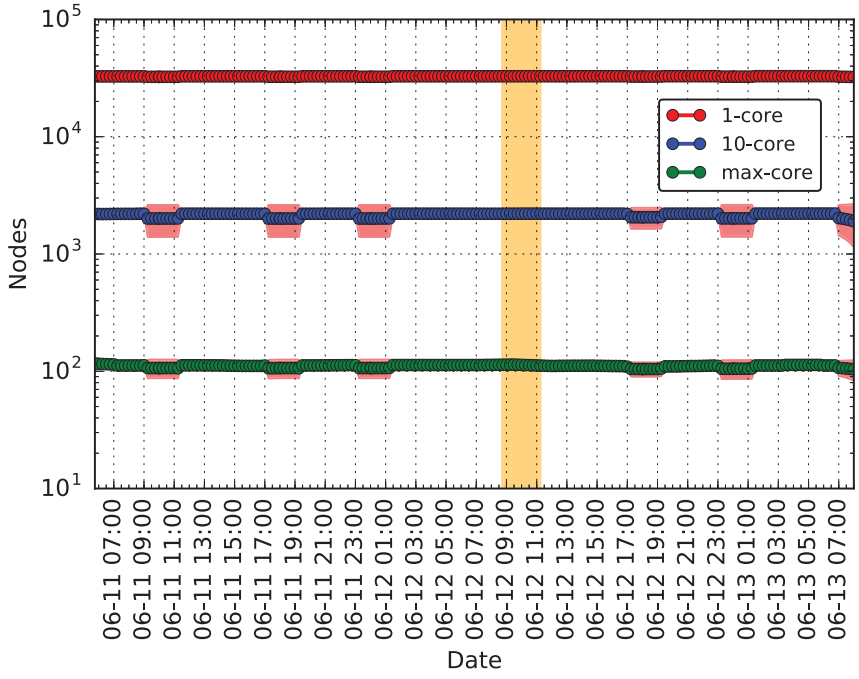

Figure A.6: Nodes per core Malaysia event.

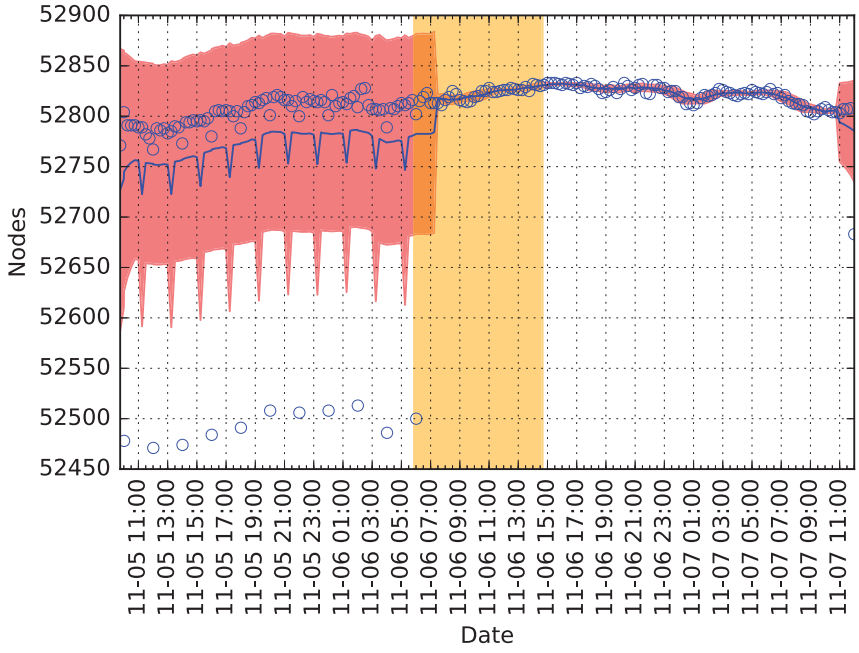

Figure A.7: Number of nodes India event.

around the same time as we observed before in the centrality and average path length measurements.

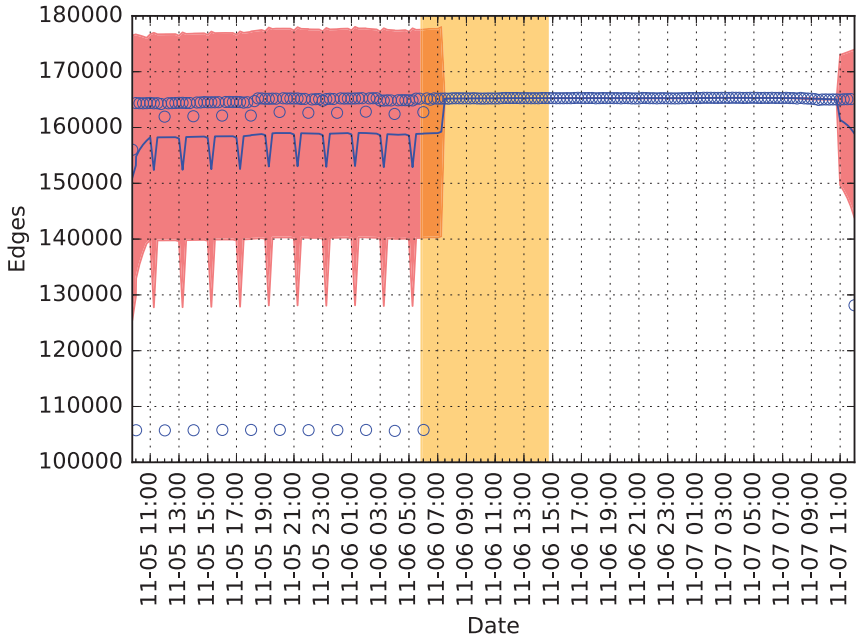

Figure A.8: Number of edges India event.

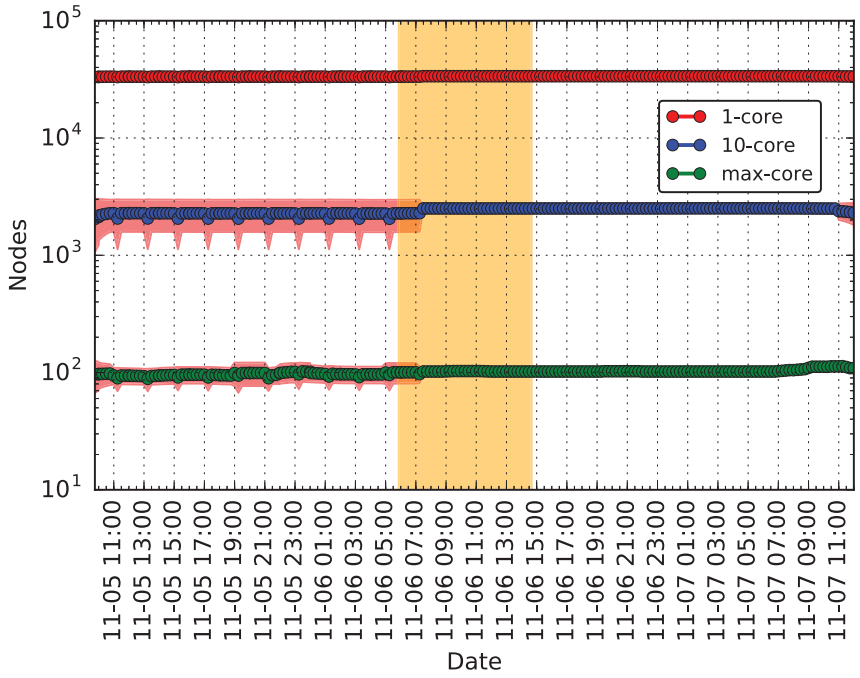

Figure A.9: Nodes per core India event.

We then looked at the average clustering coefficient in the core subgraphs for different values of k in Figure A.11. It is worth noting that—in advance—of the reported times of the incidents, it

is possible to observe some disruptions in the clustering measure for the core subgraphs.

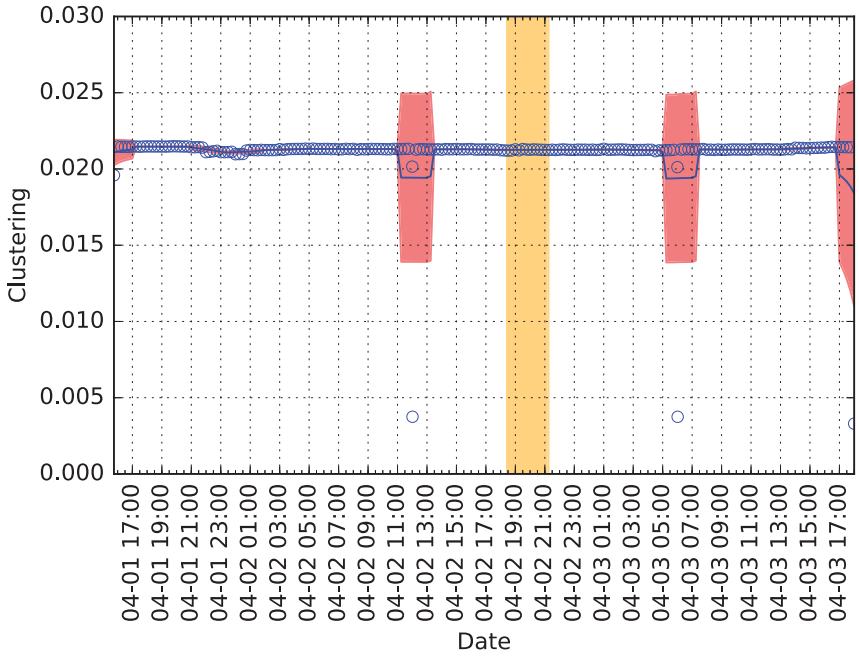

Figure A.10: Clustering coefficient Indonesia event.

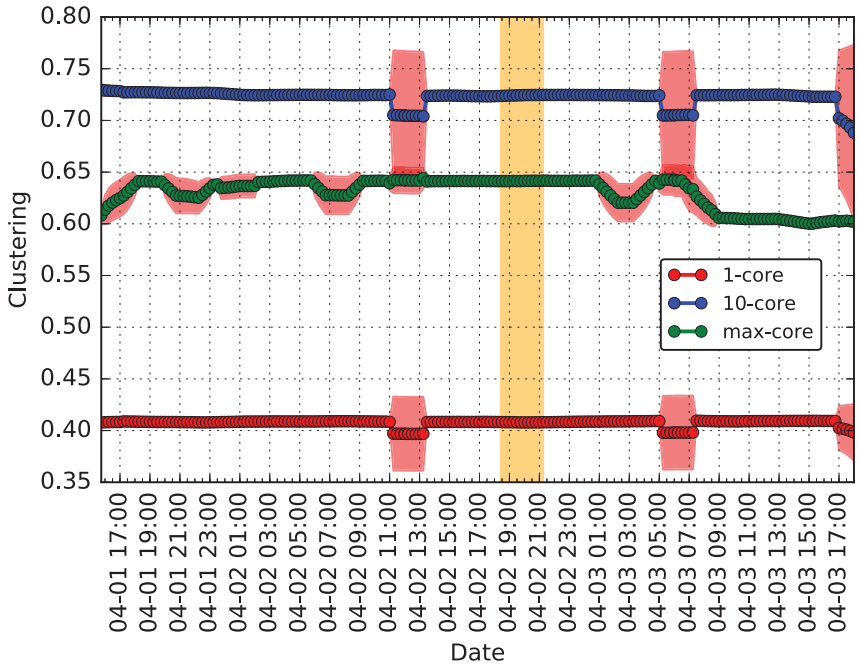

Figure A.11: Clustering per core Indonesia event.

A.2.2 Global Collateral Damage of Telecom Malaysia Leak

For the Malaysian incident, Figure A.12, shows the average clustering coefficient for the whole graph—with no k-shell decomposition applied yet. Discontinuities in the signal are observed in correspondence with the same behavior exhibited for other structural properties measured at the general graph, e.g., Figure A.4. We also studied the patterns in the number of nodes in the core subgraphs. Figure A.13 shows the variability in this pattern. It seems to coincide with previous illustrated discontinuities for the whole graph snapshots.

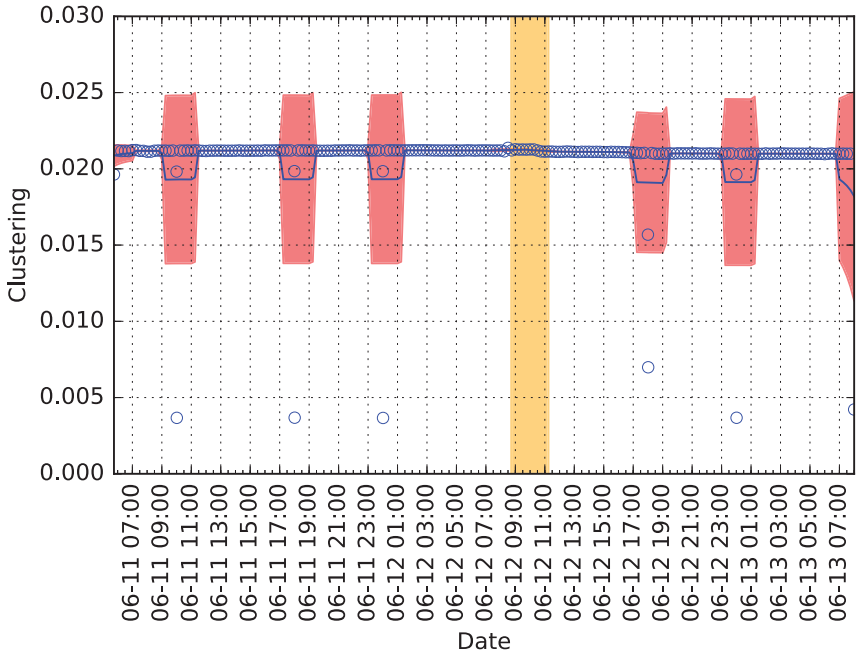

Figure A.12: Clustering coefficient Malaysia event.

A.2.3 Large Scale BGP Hijack in India

Finally, for the Indian incident, we report similar metrics in the clustering measurements as for previous anomalous events. Figure A.14 shows the time series of the average clustering coefficient. In general, the signal has discontinuities in accordance with centrality measures plots. Figure A.15 captures the same property for core and crust subgraphs. It is of interest that for core measurements,

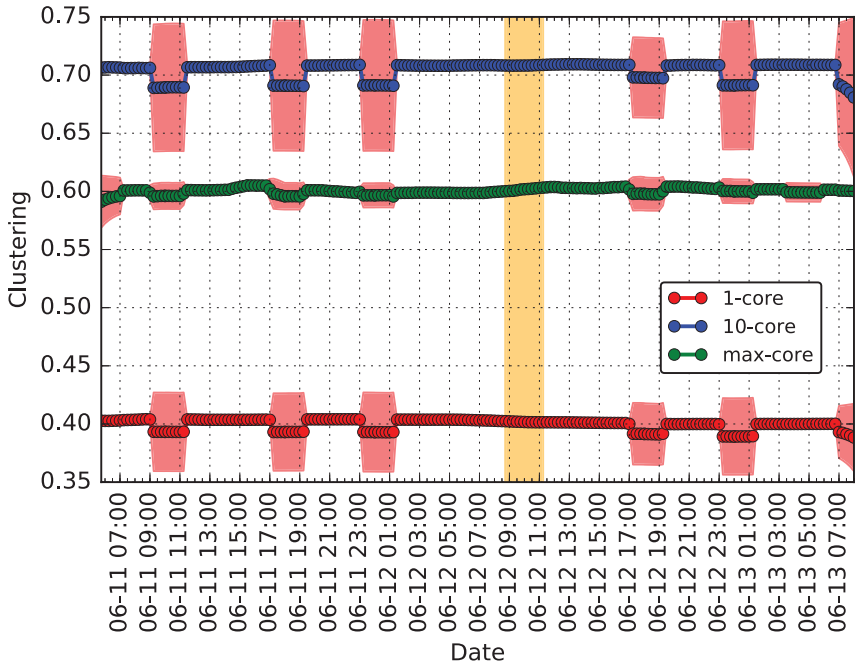

Figure A.13: Clustering per core Malaysia event.

there is a significant reduction in the clustering even under the presence of discontinuities as noticed in the case of centrality measures.

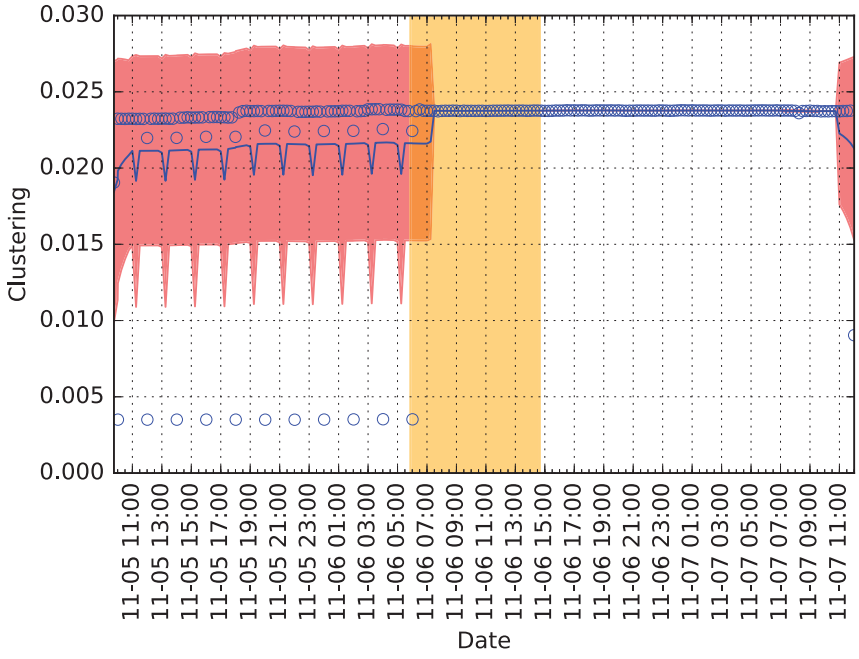

Figure A.14: Clustering coefficient India event.

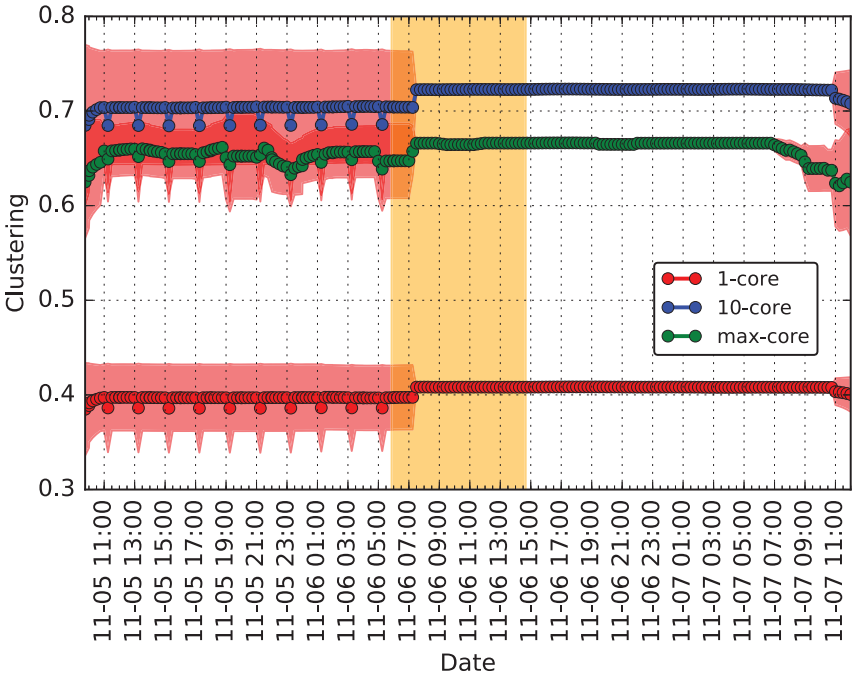

Figure A.15: Clustering per core India event.

B Remaining Collectors BGP Burstiness Analysis

B.1 Collectors' Disruption Perception

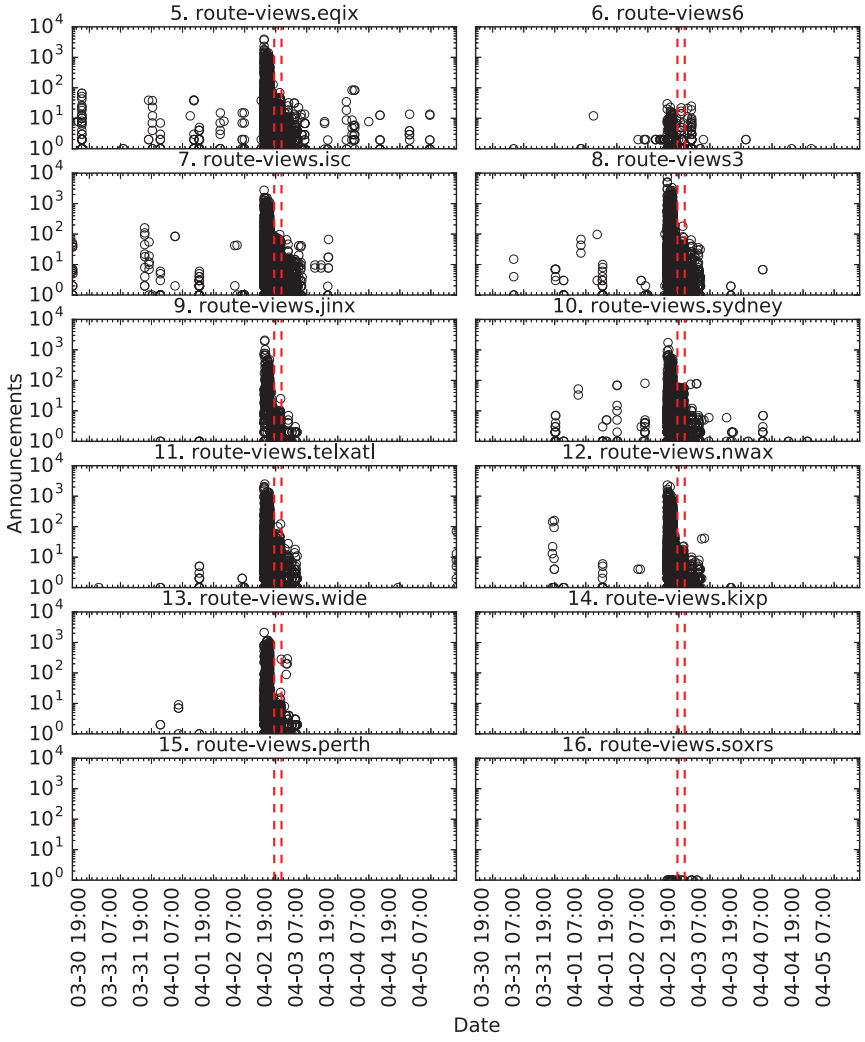

Figure B.1: Time series of the number of announcements from AS 4761 that collectors received before, during, and after the Indosat incident in 2014 for the top four collectors. Major ticks correspond to six-hour intervals while minor ticks correspond to two-hour intervals.

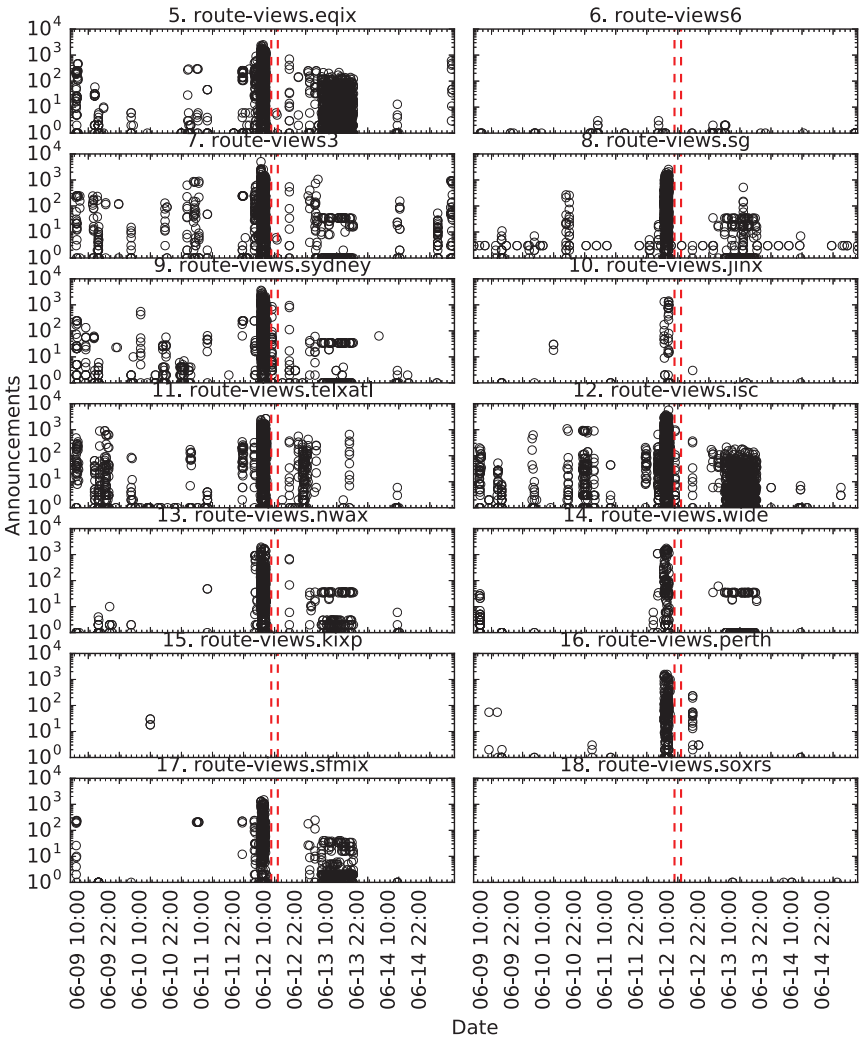

Figure B.2: Time series of the number of announcements from AS 4788 that collectors received before, during, and after the Telecom Malaysia incident in 2015.

B.2 Inter-Arrival Time Analysis

B.3 Anomaly Detection

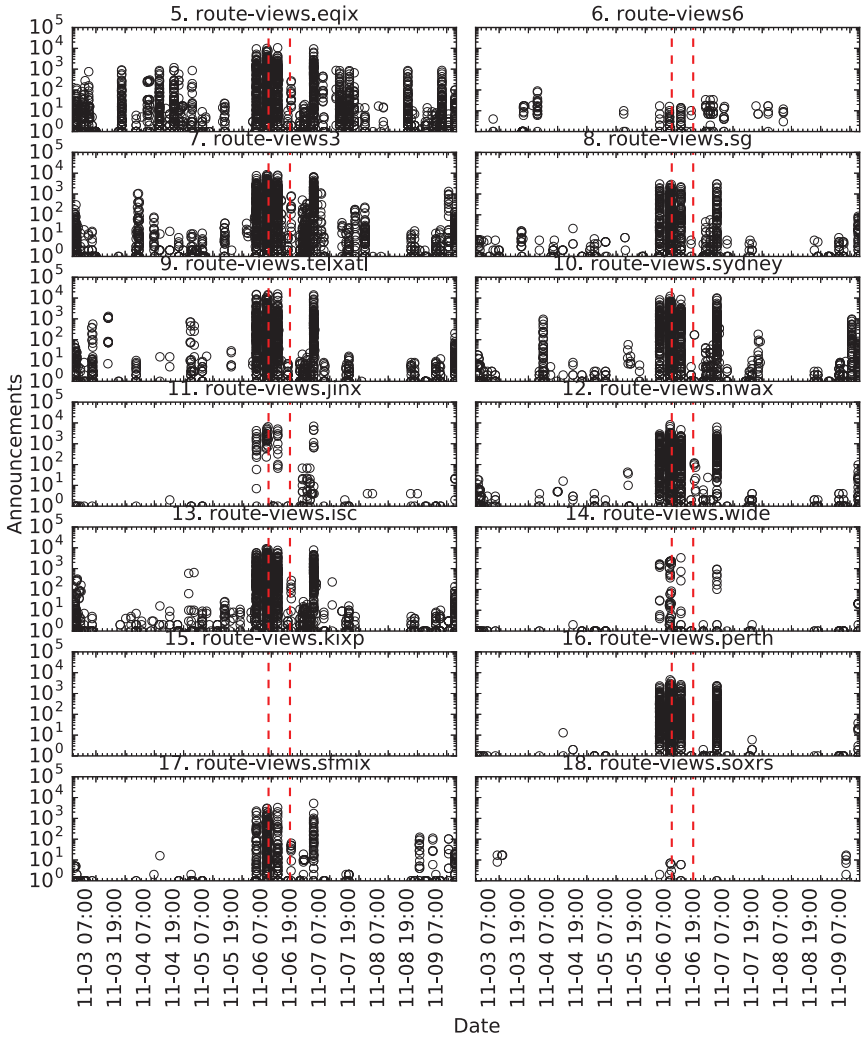

Figure B.3: Time series of the number of announcements from AS 9498 that collectors received before, during, and after the Bharti Airtel Ltd. incident in 2015.

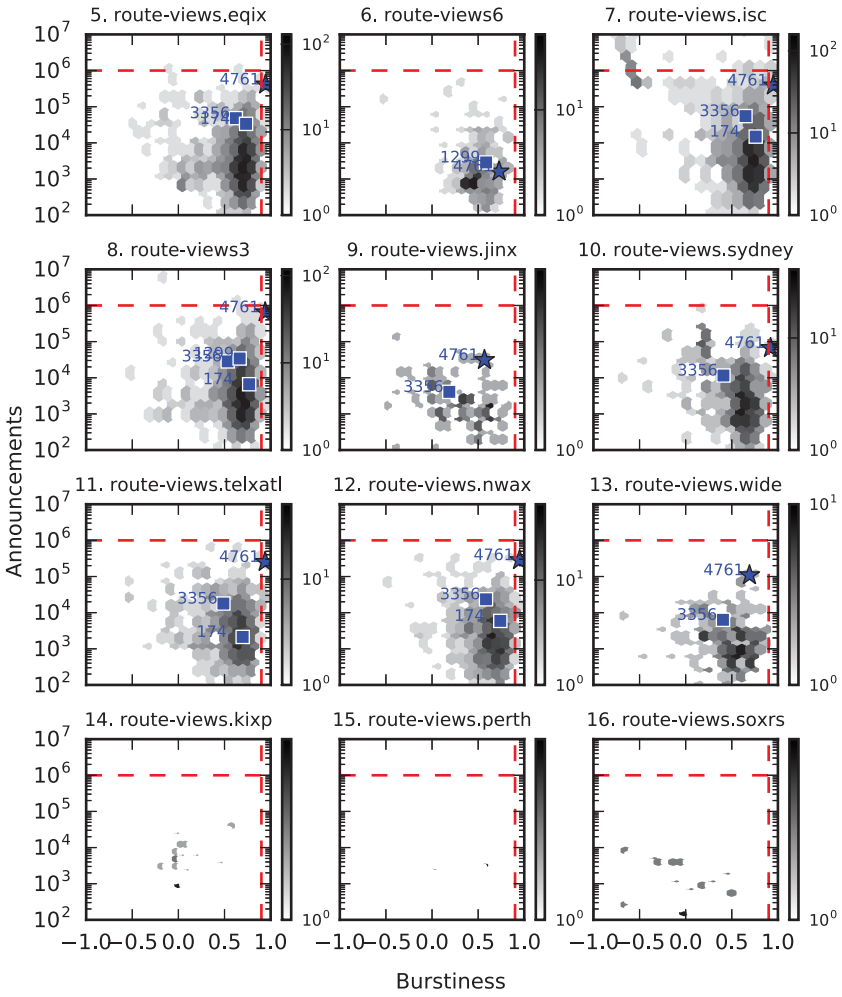

Figure B.4: Joint distribution based on the the burstiness (horizontal axis) and number of announcements (vertical axis) during one day interval around the Indosat incident.

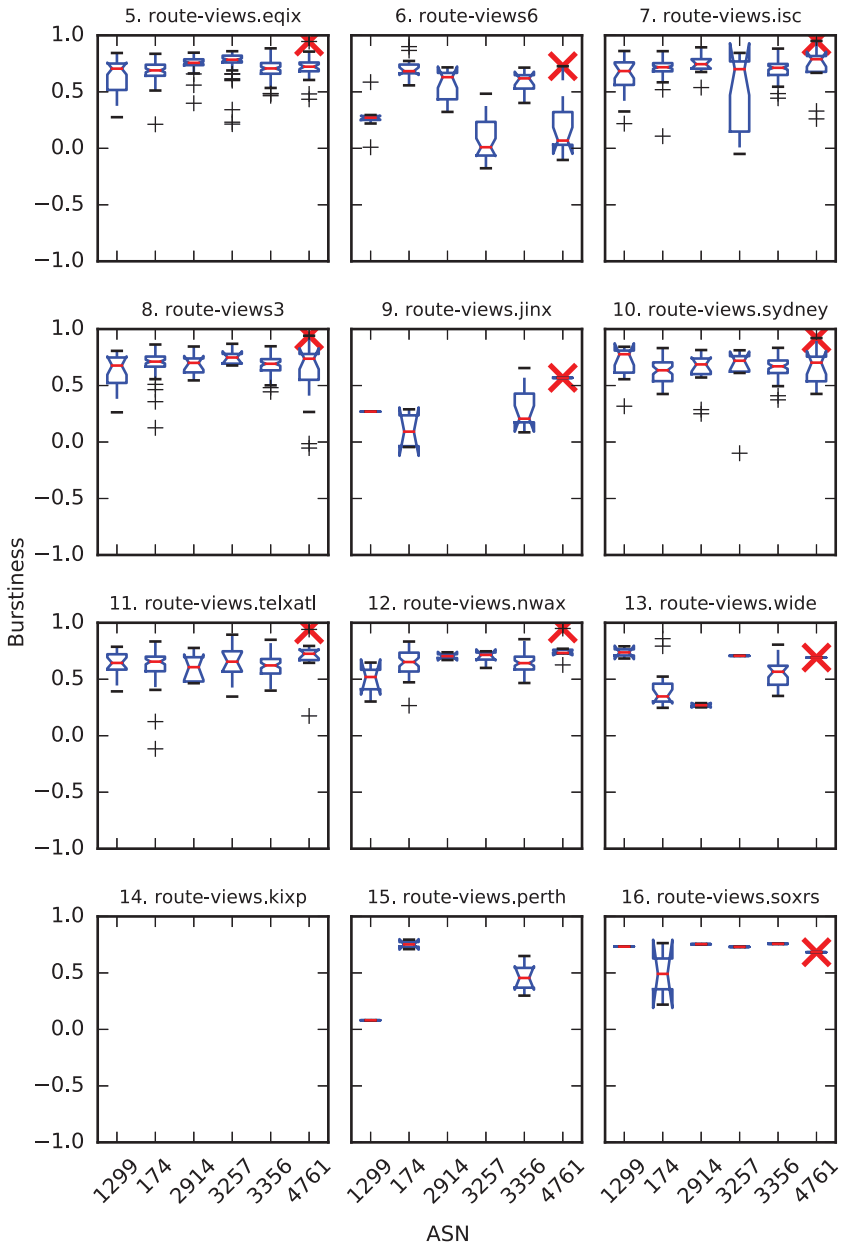

Figure B.5: Monte Carlo test for burstiness. Last column corresponds to the observations of the AS responsible for the incident, AS 4761. The test statistic, the burstiness observed during the interval of the attack, is marked with a cross.

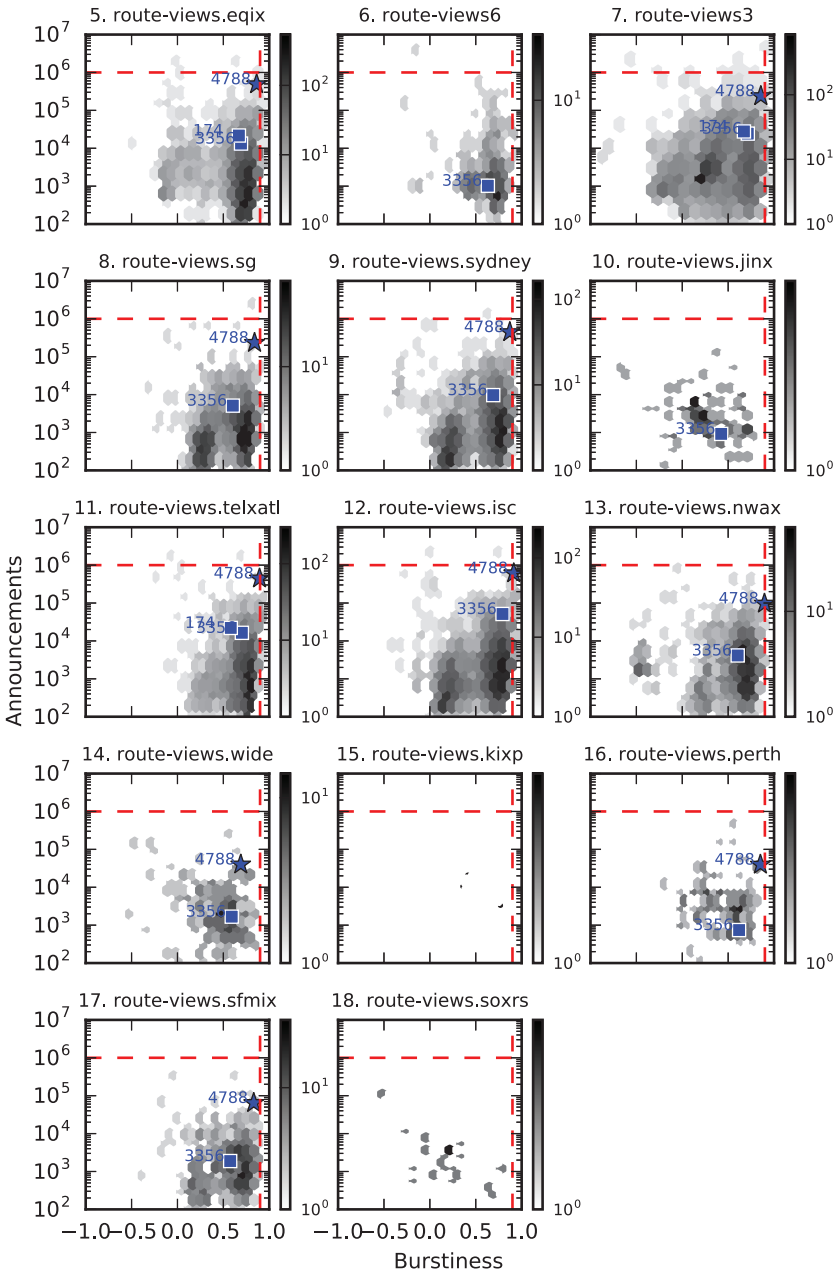

Figure B.6: Joint distribution based on the total number of announcements and their burstiness during one day interval around the Telecom Malaysia incident.

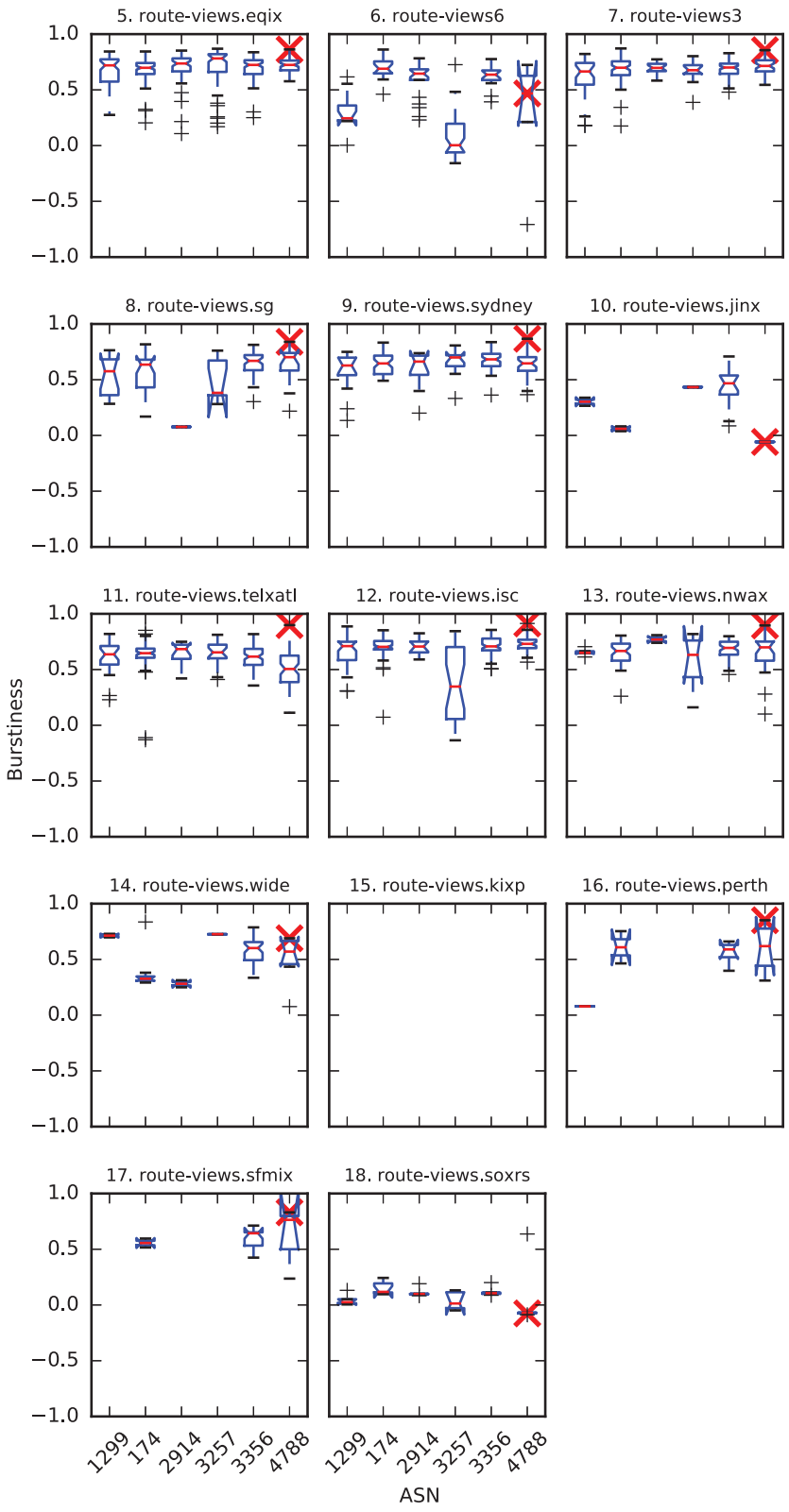

Figure B.7: Monte Carlo test for burstiness. The last column corresponds to the observations of the AS responsible for the incident, i.e., AS 4788.

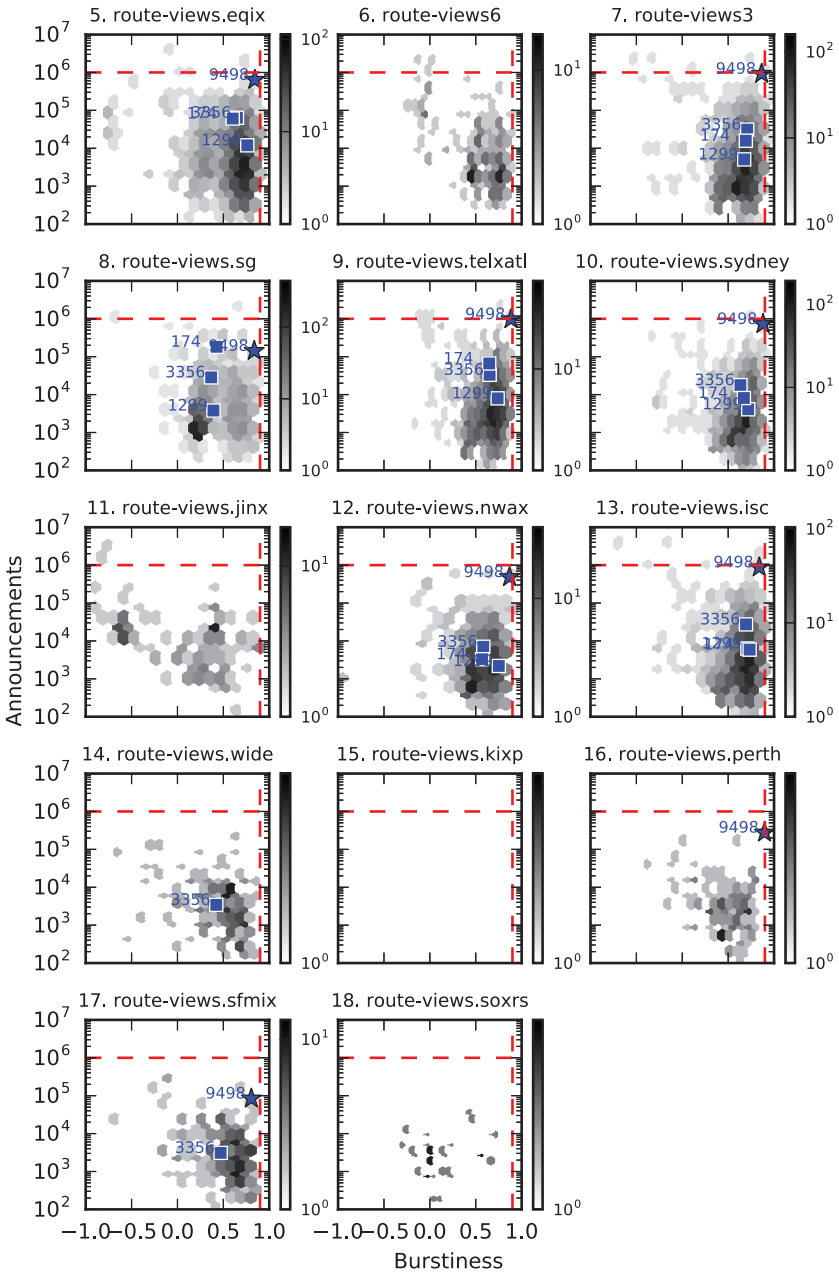

Figure B.8: Joint distribution based on the total number of announcements and their burstiness during the one day interval around the Bharti Airtel Ltd. incident.

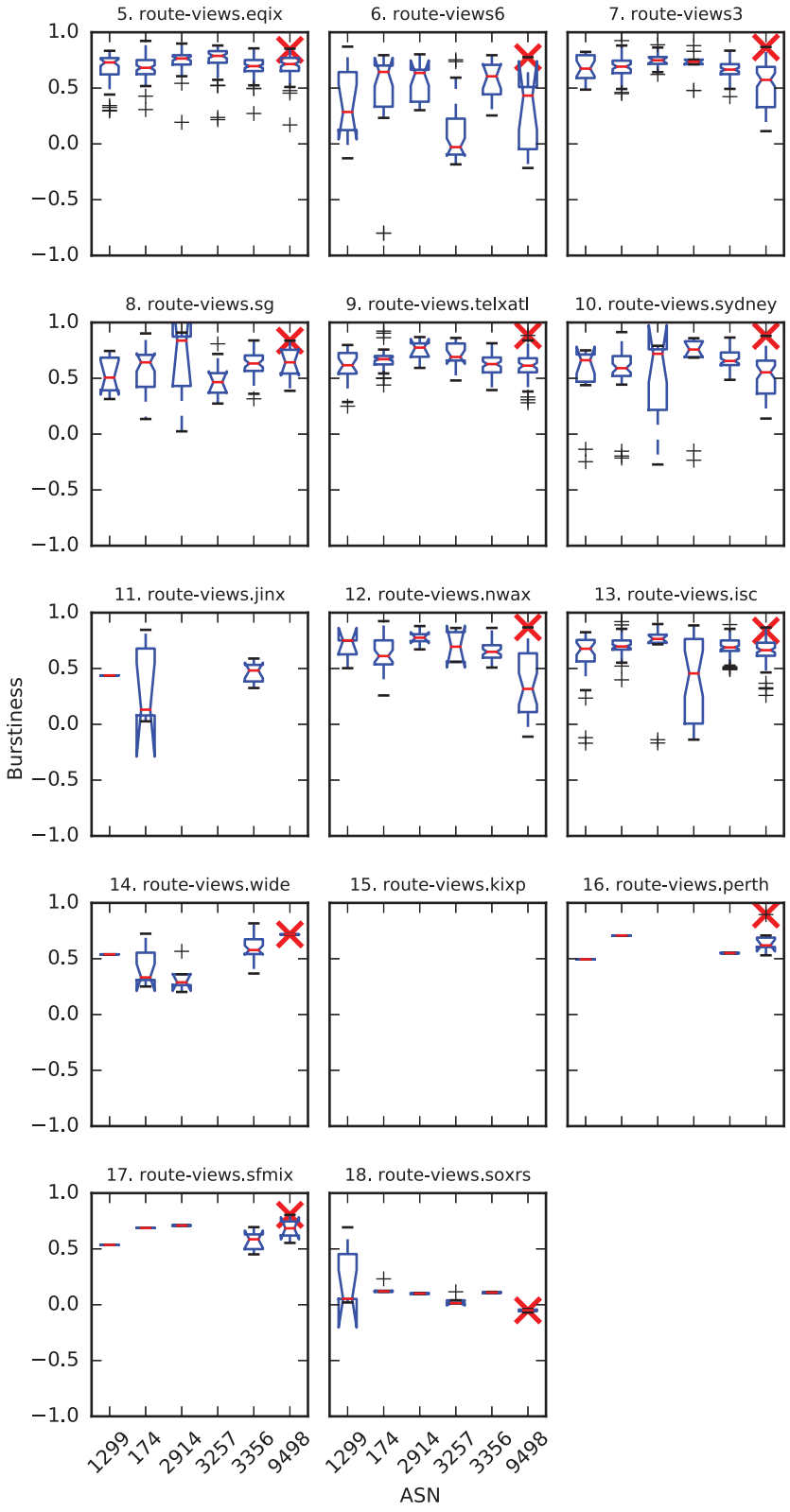

Figure B.9: Monte Carlo test for burstiness. The last column corresponds to the observations of the AS responsible for the incident, i.e., AS 9498.

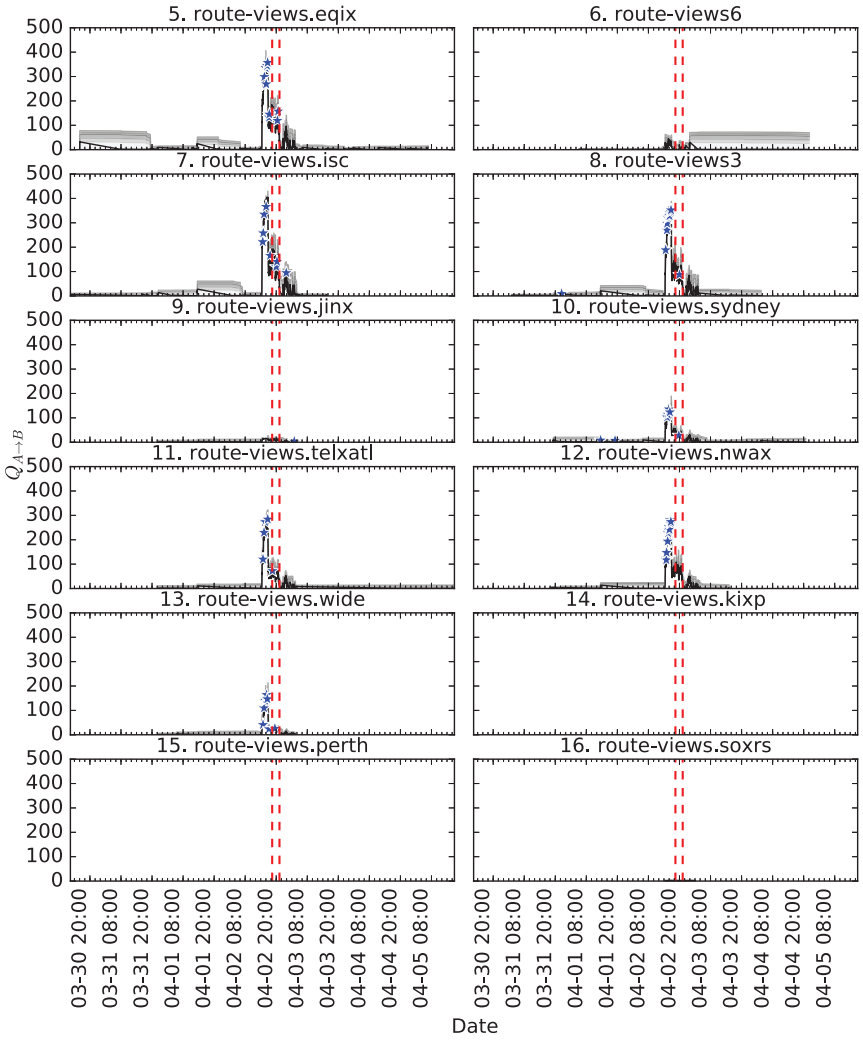

Figure B.10: $Q_{4761 \to B}$ time series for the Indosat incident.

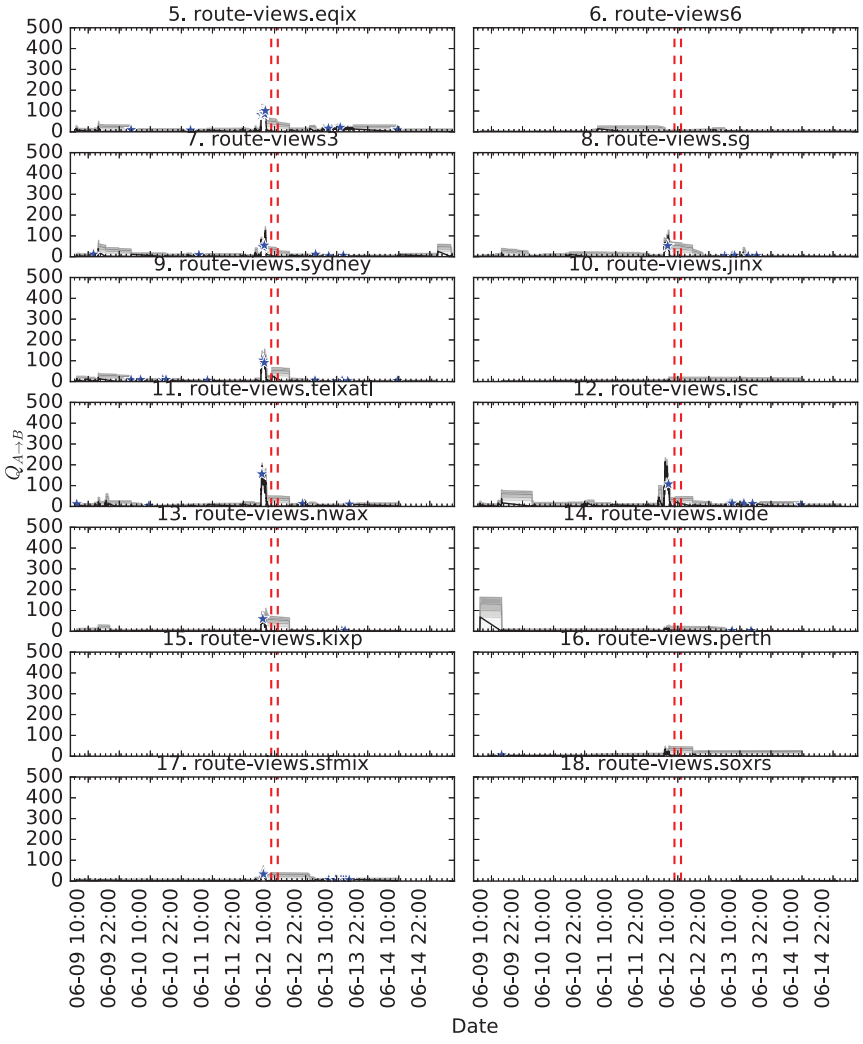

Figure B.11: $Q_{4788\to B}$ time series for the Telecom Malaysia incident.

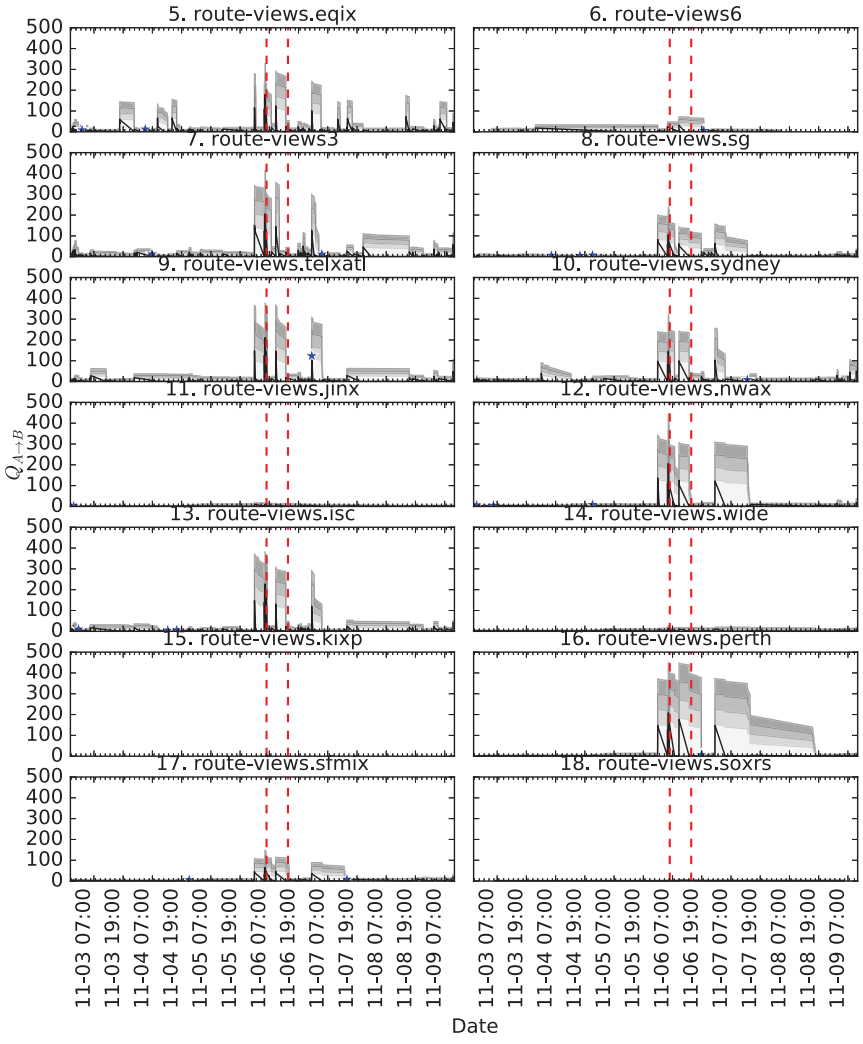

Figure B.12: $Q_{9498\to B}$ time series for Bharti Airtel Ltd.

Pablo Moriano

Research interests

Data Science, Machine Learning, Network Science, Cybersecurity, Complex Systems

Education

Indiana University, Bloomington, IN
Ph.D., Informatics, May 2019

- Dissertation: Anomaly Detection in Real-World Temporal Networks
- Committee: L. Jean Camp (Chair), Yong-Yeol Ahn, Filippo Radicchi, Raquel Hill
- Minor in Statistical Science

M.S., Informatics, October 2017

Pontificia Universidad Javeriana, Cali, Colombia
M.S., Electrical Engineering, October 2011

- Master thesis: Heavy-tailed distributions from local decision-making strategies
- Advisor: Jorge Finke
- *Summa cum laude*, with highest distinction
- Ranked top 1%, GPA: 4.74/5.00

B.S., Electrical Engineering, May 2008

- *Summa cum laude*, with highest distinction
- Ranked top 1%, GPA: 4.45/5.00

Publications

Peer Reviewed Journals

[J7] **P. Moriano**, J. Finke, and Y.-Y. Ahn. **Community-Based Event Detection in Temporal Networks**. Accepted for publication in *Scientific Reports*, 2019.
[J6] **P. Moriano**, J. Pendleton, S. Rich, and L. J. Camp. **Stopping the Insider at the Gates: Protecting Organizational Assets Through Graph Mining**. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 9, no. 1, pp. 4–29, 2018.
[J5] **P. Moriano**, S. Achar, and L. J. Camp. **Incompetents, criminals, or spies: Macroeconomic analysis of routing anomalies**. *Computers & Security*, vol. 70, pp. 319–334, 2017.
[J4] P. Rajivan, **P. Moriano**, T. Kelley, and L. J. Camp. **Factors in an end user security expertise instrument**. *Information and Computer Security*, vol. 25, no. 2, pp. 190–205, 2017.
[J3] **P. Moriano** and J. Finke. **On the formation of structure in growing networks**. *Journal of Statistical Mechanics: Theory and Experiment*, 2013 (06), P06010.
[J2] **P. Moriano** and J. Finke. **Power-law weighted networks from local attachments**. *Europhysics Letters*, vol. 99, no. 1, p.18002(6), 2012.
[J1] **P. Moriano** and F. Naranjo. **Modelado y control de un nuevo sistema bola viga con levitación magnética**. *Revista Iberoamericana de Automática e Informática Industrial*, vol. 9, no. 3, pp. 249–258, 2012.

Peer Reviewed Conferences

[C6] **P. Moriano**, R. Hill, and L. J. Camp. **Early Detection of BGP Routing Anomalies From Bursty Announcements**. *Under review*, 2019.
[C5] P. Rajivan, **P. Moriano**, T. Kelley, and L. J. Camp. **What Can Johnny Do? – Factors in an End-User Expertise Instrument**. In Proceedings of the *Tenth International Symposium on Human Aspects of Information Security & Assurance (HAISA)*, pp. 199–208, Frankfurt, Germany, July 2016.

[C4] **P. Moriano** and J. Finke. **Model-based fraud detection in growing networks**. In Proceedings of the *IEEE Conference on Decision and Control (CDC)*, pp. 6068–6073, Los Angeles, CA, USA, December 2014.
[C3] **P. Moriano** and J. Finke. **Characterizing the relationship between degree distributions and community structures**. In Proceedings of the *American Control Conference (ACC)*, pp. 2383–2388, Portland, OR, USA, June 2014.
[C2] **P. Moriano** and J. Finke. **Structure of growing networks with no preferential attachment**. In Proceedings of the *American Control Conference (ACC)*, pp. 1088–1093, Washington, DC, USA, June 2013.
[C1] **P. Moriano** and J. Finke. **Heavy-tailed weighted networks from local attachment strategies**. In Proceedings of the *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, pp. 5211–5216, Orlando, FL, USA, December 2011.

Referred Workshops

[W3] **P. Moriano**, J. Pendleton, S. Rich, and L. J. Camp. **Insider Threat Event Detection in User-System Interactions**. In Proceedings of the *9th ACM CCS International Workshop on Managing Insider Security Threats (MIST)*, pp. 1–12, Dallas, TX, USA, October 2017 (**Best paper award**).
[W2] **P. Moriano**, E. Ferrara, A. Flammini, and F. Menczer. **Dissemination of scholarly literature in social media**. In Proceedings of the *ACM Web of Science Conference Workshop Altmetrics*, Bloomington, IN, USA, June 2014.
[W1] **P. Moriano** and F. Naranjo. **Modelado de un nuevo sistema bola viga con levitación magnética**. In Proceedings of the *4th IEEE Colombian Workshop on Robotics and Automation*, Cali, Colombia, August 2008.

Referred Abstracts & **Posters**

[A4] **P. Moriano**, R. Hill, and L. J. Camp. **Hijacking Network Traffic: Temporal Analysis of Adverse Changes in the Internet Topology**. In *Conference on Complex Systems (CCS)*, Thessaloniki, Greece, September 2018.
[A3] C. McElroy, **P. Moriano**, and L. J. Camp. **On Predicting BGP Anomalous Incidents: A Bayesian Approach**. In *Network and Distributed Security Symposium (NDSS)*, San Diego, CA, USA, February 2018 (**Honorable mention**).
[A2] **P. Moriano**, J. Finke, and Y.-Y. Ahn. **Community-based anomalous event detection in temporal networks**. In *Conference on Complex Systems (CCS)*, Cancún, Mexico, September 2017.
[A1] **P. Moriano**, S. Achar, and L. J. Camp. **Macroeconomic Analysis of Routing Anomalies**. In *Telecommunications Policy Research Conference (TPRC)*, Arlington, VA, USA, October 2016 (**Honorable mention**).

Work in Progress

[WP1] **P. Moriano**, C. McElroy, and L. J. Camp. **On predicting BGP hijacking incidents: A Bayesian approach**. In preparation.

Research Experience

Indiana University, Bloomington, IN
Research Assistant **June 2015 to Present**

- Analyzed a dataset of routing anomalies using unsupervised machine learning methods to understand country-based generation of those.
- Collected a dataset of BGP routing updates for time series analysis of hijacking events.

- Conducted network analysis on BGP updates and proposed a framework of early identification of large-scale network disruptions.
- Performed statistical analysis of large-scale computer security surveys to distinguish traits between experts and non-experts security practitioners.
- Published 3 first author research articles on data-driven security applied to routing anomaly detection.
- Devised projects while teaching and mentoring 1 undergraduate and 3 graduate students.

PI: **L. Jean Camp**

Research Assistant **September 2013 to July 2014**

- Conducted Twitter data analysis to understand how scientific publications spread online and presented results at an international conference.

PIs: **Filippo Menczer** and **Alessandro Flammini**

Cisco Systems, Inc., Knoxville, TN

Research Intern **Summers 2016, 2017, and 2018**

- Designed and implemented an anomaly detection method based on temporal network analysis for identifying suspicious commits in Cisco's IOS codebase.
- Established collaborations to conduct experiments requiring specific techniques.
- Published a first author research article on insider threat event detection in the 9th ACM CCS International Workshop on Managing Insider Security Threats (MIST), which results in best paper award.
- Presented results at an international conference attended by more than 500 scientists.
- Participated in additional research that lead to an accepted research proposal for investigating vulnerability prediction in Cisco's codebases for over $30,000.
- Reported progress at regular meetings with the company SVP.

Mentor: **Steven Rich**

Pontificia Universidad Javeriana, Cali, Colombia

Research Assistant **February 2009 to July 2013**

- Developed software for constructing models of networks that have both heavy-tail degree distributions and high degrees of clustering.
- Participated in additional research that lead to an accepted research proposal with Colombian's National Science Department for investigating methods for anomaly detection in networks for $10,000.
- Published 3 first author research articles on mechanisms of network formation.
- Presented results at 3 international conferences in control systems.

PI: **Jorge Finke**

Teaching Experience

Indiana University, Bloomington, IN

Associate Instructor **August 2014 to May 2015**

- Assisted in teaching 2 undergraduate courses ranging in size from 20-80 students on topics including: Discrete mathematics, programming in Python, and statistics.
- Led weekly laboratory and/or problem-solving and discussion sections for groups of 5-10 students.
- Supervised students in final projects, graded exams and weekly homework.

Pontificia Universidad Javeriana, Cali, Colombia

Lecturer **July 2011 to July 2013**

- Recognized as an outstanding lecturer while teaching an undergraduate introduction to programming class of about 30 students.

- Prepared course material including laboratory experiments, lectures, exams, homework, and practice problems.

Awards

Cisco Systems, Inc.
- PI: Understanding Software Quality in Developer-Component Temporal Graphs, May 2018 ($32,000)

Indiana University, Bloomington, IN
- Research and teaching assistantship, 2013–2017

Colciencias, Bogotá, Colombia
- Science, technology, and innovation scholar, January 2014
- Young researcher award from Colombia's National Science Agency, October 2010 ($10,000)

Colfuturo, Bogotá, Colombia
- Graduate student scholarship, June 2013

Pontificia Universidad Javeriana, Cali, Colombia
- Outstanding lecturer, May 2013
- Outstanding master thesis, October 2011
- M.S. research scholarship, 2009–2011
- Outstanding undergraduate thesis, May 2008
- Dean's List, 2003–2007

Travel Grants
- CRA Grad Cohort Workshop for URMD, 2019
- Tapia Conference Doctoral Consortium, 2018 ($1,500)
- IU Graduate and Professional Student Government, 2017 ($500)
- IEEE Symposium on Security and Privacy (IEEE S&P), 2017 ($900)
- GREPSEC III Workshop (supported by NSF), 2017 ($700)
- American Control Conference (ACC), 2014 ($800)

Best Paper Award
- 9th ACM CCS International Workshop on Managing Insider Security Threats (MIST), 2017

Community service

Memberships
- Institute of Electrical and Electronics Engineering (IEEE) student member
- Association for Computing Machinery (ACM) student member
- Complex Systems Society (CSS) member
- Federation of Automatic Control (IFAC) technical committee member for Technology, Culture, and International Stability

Mentoring and Advising
- Clint McElroy, B.S. in Informatics, Indiana University, 2017-2018
- Srivatsan Iyer, M.S. in Computer Science, Indiana University, 2015-2017
- Soumya Achar, M.S. in Computer Science, Indiana University, 2015-2016

Master Thesis Reviewer
- Juan Camilo Campos, M.S. in Electrical Engineering, Pontificia Universidad Javeriana, 2018

Reviewing

Journal Referee
- IEEE Transactions on Knowledge and Data Engineering
- IEEE Access
- ACM Transactions on Information and System Security (TISSEC)

Technical Program Committees
- ACM Internet Measurement Conference (Shadow PC 2017)

Skills

Programming Languages

- Frequent user of Python for data analysis using Matplotlib, igraph, Pandas, Scikit-learn
- Experience in LaTeX, R, MATLAB, Mathematica, C/C++
- Familiar with HTML, CSS, JS for frontend
- Used SQLite, NoSQL (MongoDB)

Spoken Langauges

- English (fluent), Spanish (native)

Extracurricular Activities

- Tennis, travel, hiking